Ecology, Virulence and Genomics of *Legionella pneumophila* Isolates from the West Bank, Palestine

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Ashraf R.N. Zayed

aus Nablus/Palästina

1. Referent: apl.Professor Dr. Manfred G. Höfle

2. Referent: Professor Dr. Michael Steinert

eingereicht am: 17.08.2018

mündliche Prüfung (Disputation) am: 30.10.2018

Druckjahr 2018

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Legionella in hospital water of the West Bank: Analysis of environmental drivers and virulence assessment of relevant MLVA genotypes. Ashraf R. Zayed, Lina Abu Tair, Suha Butmeh, Shereen A. Bahader, Marina Pecellin, Ingrid Brettar, Manfred G. Höfle, and Dina M. Bitar. International symposium on *Legionella* – Ecology, Virulence and Risk Assessment, September 2016, Braunschweig, Germany. (Oral presentation)

Acknowledgements

First and foremost, I thank Allah for giving me the strength to persevere through this endeavor

I would like to express my deepest gratitude and appreciation to every person who contributed and made this research work possible and to the **German Research Foundation (DFG)** for funding the project.

My deepest gratitude, appreciation and respect to my mentor **Prof. Dr. Manfred Höfle** and my supervisors **Dr. Dina M. Bitar** and specially **Dr. Ingrid Brettar** for giving me the opportunity to join their research project, for their enthusiasm, inspiration ,great efforts to explain things clearly and simply. I am grateful for their endurance. Editing suggestions and precise sense of language, which contributed to the final copy of my dissertation.

I would like to thank all my **MIDI group** colleagues: **Josi, Marina, Douglas, Rui** and specially **Verena** and **René** for their support, interest and valuable hints.

I would like to express my deeply-felt and sincere gratitude to **Al-Quds University** represented by all the teachers and colleagues. I am deeply thankful for all my AQU colleagues: **Dr. Lina Abu Tair, Hanna Alalam, Alaa Salah, Dalia Abu Hilal and Dr. Shereen Bahader**. My appreciation is extended to **Suha Butmeh** for her excellent assistance, comments in the lab, support and friendship and for all she did for me.

Let me thank and appreciate **Mr. Omar Zayed** from the Palestinian Water Authority (PWA) who provided part of the physico-chemical parameters and valuable references regarding the water status in the West Bank.

I am deeply indebted to the super Bioinformatics specialist **Dr. Boyke Bunk** from DSMZ. He taught me a lot of new things, which I really like. Also, let me thank and appreciate **Prof. Dr. Michael Steinert** for kindly accepting to be the second reviewer of my dissertation and for his time and comments during thesis committee discussions.

Also, big thanks to several friends with whom I spent good times at the HZI; **Ahmad, Aida Alessandro, Chandra, Dario, Kamila, Kihn, Mohammad Hasan, Mumdooh, Omar, Siti, Soleiman** and **Vivian** and outside the HZI; **Abed-Allah, Ameni, Arwa, Ghanem, Mohammad Kamel, Mohammad Salhi, Nasser, Saber and Saher.** My thanks are extended to my best friend **Yassin**. Thank you for everything “Ya Zoul”

Finally, I wish to thank with tearful heart **my** great **father Rashad** whose time on earth came and went before my eyes, who left this life so fast and quietly as he has always been. I got my biggest inspiration to do this PhD from **you**. Great thanks to my dear brother, **Majd**, my dear wife **Aseel** and my beautiful kids **Rashad** and **Laila**. Last but foremost, my very great appreciation to **my** great **mother Hana**, I am so lucky to have a supportive mother like you who listens to me and to whom I can open up my world to. No words can explain how grateful I am to have a mother like you.

Summary

Legionella pneumophila is an environmental bacterium and a human pathogen causing life-threating outbreaks of an atypical pneumonia called Legionnaires' disease (LD). Studies of this pathogen have focused mainly on Europe and the United States. A shortage in *L. pneumophila* data is clearly observed for developing countries. To reduce this knowledge gap, *L. pneumophila* isolates were studied in two widely different geographical and climatic areas, i. e. Palestine and Germany, with emphasis on Palestine for the presented thesis. In addition, this thesis aimed to understand the diversity of *L. pneumophila* isolates, their clonal populations and the environmental driver of their abundance and prevalence.

For this purpose, a multi-annual seasonal sampling of nine different drinking water sites in the West Bank were performed including a detailed recording of the respective physico-chemical and microbiological parameters. Drinking water and its biofilm were sampled in eight hospitals and the Al-Quds University, providing an overview of the West Bank. A total of 180 *L. pneumophila* isolates were obtained and subsequently analysed by high resolution genotyping (Multi-Locus Variable Number of Tandem Repeats (MLVA-8(12) using 13 loci) in addition to standard characterization by 16S rRNA and serogroup. In addition, physiological and virulence traits were studied. Genotyping and the studied traits led to the selection of representative strains submitted to high through-put genome sequencing (Illumina HiSeq and Pacific Biosciences platforms). Analysis of genotype prevalence in correspondance to environmental factors was used to elucidate genotype consortia and their environmental niches.

The one hundred eighty isolates represented twenty-six individual MLVA-8(12) genotypes (Gt). The most frequently isolated genotype was Gt4(17) (41.1%) and Gt6(18) (16.7%), both ST1, and Gt10(93) (9%, ST461). All MLVA genotypes were clustered into four MLVA clonal complexes (VACCs). The genotype composition showed a regional variability. Analysis of the prevalence of genotype in the light of environmental factors indicated genotype consortia that seemed to be triggered by a set of environmental drivers. The concentration of several ions (Mg, Ca, Cl, SO4 and TDS) and turbidity seemed to determine niches for three different sets of genotypes and may thus explain their regional variability. To a certain extent, also the abundance of *L. pneumophila* was influenced by these environmental drivers with Mg having a negative effect. The virulence of a representative subset of sixty Palestinian environmental strains was assessed by five different in-vitro assays Virulence traits were shown to be genotype dependent, with high virulence observed for the highly abundant genotypes Gt4(17) and Gt6(18).

A carefully selected subset of thirty-eight *L. pneumophila* isolates were genome sequenced (HiSeq) and compared to already published reference genomes, and two additionally Pac-Bio sequenced isolates providing needed complete reference genomes. Genome sequences from the thirty-eight isolates were aligned with the sequence of the respective reference genome and analysed with respect to core-single nucleotide polymorphisms (core-SNPs), genomic islands and genes related to virulence traits. Overall, the results obtained from this study provided important insights into detailed population structure, the ecology and pathogenicity of this pathogen in the West Bank.

Table of Contents

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of Variance
AQU	Al-Quds University
BCYE	Buffered Charcoal Yeast Extract
Bp	Base Pairs
BLASTp	Protein Basic Local Alignment Search Tool
CBC	Complete Blood Count
CDA	Cultivation Dependent Analysis (Culturing)
CFU	Colony Forming Unit
cgMLST	core genome-Multi Locus Sequence Typing
CIA	Cultivation Independent Analysis (PCR)
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CV	Coefficient of Variation
Dot/Icm	Defective organelle trafficking/Intracellular multiplication
DNA	Deoxyribonucleic Acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DWSS	Drinking Water Supply System
ECDC	European Centre for Disease Prevention and Control
EU	European Union
EWGLI	European Working Group for *Legionella* Infection
FBS	Fetal Bovine Serum
GI	Genomic Island
GVPC	Glycine Vancomycin Polymyxin Cycloheximide *Legionella* medium
HPC	Heterotrophic Plat Count

HZI	Helmholtz Zentrum für Infektionsforschung
LCB	Locally Collinear Blocks
LD	Legionnaires' disease
MLVA	Multi-Locus Variable number of tandem repeat Analysis
MOH	Minstery of Health
MOI	Multiplicity of Infection
PacBio	Pacific Biosciences
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PE	Post Exponential
PGYE	Peptone Glucose Yeast Extract
PWA	Palestinian Water Authority
RFLP	Restriction Fragment Length Polymorphism
RPMI	Roswell Park Memorial Institute cell culture medium
rRNA	Ribosomal Ribonucleic Acid
SBT	Sequence Based Typing
SD	Standard Deviation
Sg	Serogroup
SIMPROF	Similarity Profile Analysis
SMRT	Single Molecular Real-Time
SNP	Single Nucleotide Polymorphism
sRBCs	Sheep Red Blood Cells
ST	Sequence Type
UAE	United Arab Emirates
USA	United States of America
VACC	Variable Analysis Clonal Complex

VBNC	Viable But Not Culturable
VNTR	Variable Number of Tandem Repeats
YEB	Yeast Extract Broth
WGS	Whole Genome Sequencing

1. Introduction

1.1 Drinking water in the West Bank, Palestine

The West Bank is landlocked territory located near to the Eastern coast of the Mediterranean Sea. It has land area of 5640km^2 and 220km^2 water area of the Dead Sea. It has Eastern boundaries across the Jordan valley with Jordan and surrounded by Israel from the other boundaries **(Figure 1)**(1).

In the West Bank, the Palestinian Water Authority (PWA) is responsible and coordinates with Mekorot (The Israeli Water Company) the quality of drinking water. The PWA is analyzing microbial contaminants (total and fecal coliforms), chemical contaminants (inorganic contaminants, heavy metals and ions) and physic-chemical parameters (temperature, pH and conductivity).

For the present thesis the drinking water of the West Bank was analyzed. The West Bank is supported by the PWA, which delivers drinking water to every city, where it is transferred to the municipal distribution network. The PWA provides a total of population 2.8 million inhabitants (1) with domestic water that has an average consumption of 70 L/day per person. This amount of water is 30 L/day less than WHO recommended amount (100 L/day) of water per person. Thus, Palestinians store water in collection tanks on their house's roof for weeks or months because there is no continuous water supply (2).

Ground water is the main source of raw water in the West Bank. More than 90% of the population is provided with drinking water coming from ground water. The main aquifers can be divided into three units: 1) the Western aquifer basin, 2) the North-Eastern aquifer basin and 3) the Eastern aquifer basin (2, 3). Water is pumped from deep wells to 5000m^3 tanks (balance tank) built on high altitude land to store raw water and provide pressure for the drinking water supply system (DWSS). Chlorination is only treatment process. The chlorine concentration is 0.5mg/l in water stemming from balance tanks (3).

Figure 1: The West Bank map

1.2 *Legionella pneumophila* - a natural water-based pathogen

In July 1976, an unknown bacterium causing a common source outbreak of pneumonia in Philadelphia, Pennsylvania USA, was identified. Approximately, 15% of the cases were fatal (4). By December 1976, Joseph McDade and coworkers' isolated bacterium named it *Legionella pneumophila* (5). *L. pneumophila* is a gram negative, aerobic, rod shape flagellated bacterium that is naturally found in freshwater environments. Until now, there are 59 distinct *Legionella* species encompassing 70 serogroups or more. Approximately, half of these serogroups have been isolated from or detected in clinical samples. All these species are considered as potential human pathogens and their abundances should be controlled and/or regulated (6, 7).

1.2.1 Ecology of *Legionella*

Legionella spp are ubiquitous in natural freshwater habitats like rivers, lakes and hot springs. Naturally, species of *Legionella* are part of the freshwater microbiome (8). In these environments, *Legionella* species can be associated with complex biofilm communities which provide shelter, nutrients and support their survival and multiplication (9, 10). Also, *Legionella* have the ability to infect and replicate inside amoeba (11). The unique physiology of *Legionella* is adapted for survival and replication within amoeba, such as *Acanthamoeba,* (12, 13) and secondarily as free-living or biofilm-associated bacteria. The association with amoeba provides *Legionella* with different characteristics including induction of virulent phenotype, provision of protection from harsh conditions and support in distribution throughout the DWSS (14-16). *Legionella* thrive in freshwater ideally at 25-37°C but *Legionella* can survive at temperature below 20°C and above 55°C (17, 18). Since 1976, all *Legionella* studies demonstrates the adaptability and colonization in man-made aquatic environments mainly hospitals and hotels (19, 20).

1.2.2 Virulence of *Legionella*

Legionella spp have developed mechanisms for invasion and multiplication in protozoan hosts in the course of evolution. The same mechanisms of intracellular invasion that they use when multiplying in *amoeba* seem to operate in alveolar macrophages (21, 22). When amoeba or macrophages have ingested virulent *Legionella* cells, a phagosome is established; this is surrounded by endoplasmic reticulum and becomes completely isolated from the endosomal pathway (21, 22). Initially, fusion with lysosomes is inhibited. It has been postulated that *L. pneumophila* converts to a replicate form in this protected environment (22). Accordingly, endosomes containing the pathogen are able to fuse with lysosomes enabling the intracellular bacteria to make use of a nutrient-rich niche, which in ordinary circumstances would kill other bacteria. When the amino acid supply is depleted, the cells convert to a stationary phase form, simultaneously developing features that are needed for transmission to a new phagocyte. *Legionella* cells released from eukaryotic cells are short, thick, and highly motile. Thus, they exist in nature in two phases. In this system, a number of factors are implicated including type II and IV secretion system, acquisition of iron, pore-forming toxins, and induction of apoptosis in the host cell. The intracellular

establishment and trafficking of *Legionella* are believed to be regulated by the *dot/icm* (defective organelle trafficking/intracellular multiplication) gene complex, which encodes the compounds involved in the type IV secretion system **(Figure. 2)**. Also, *Legionella* produce extracellular cytotoxins. Experimental work indicates that virulence is significantly reduced when the incubation temperature of a cultured inoculum is reduced from 37°C to 24°C (23).

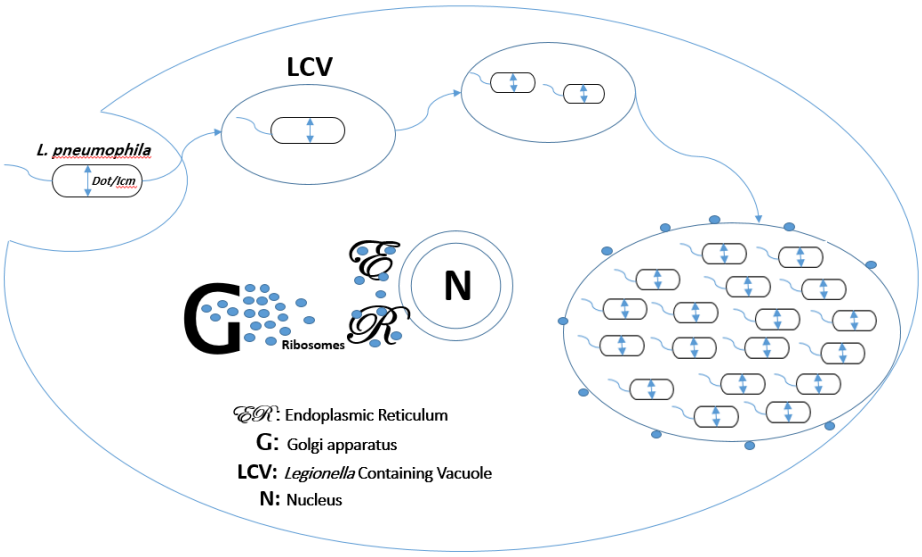


Figure 2: *L. pneumophila* infection of an amoeba/macrophage using the Dot/Icm secretion system

Legionella infect humans by inhalation of contaminated aerosols (22, 24). This infection route is supported by the fact that patients in case control studies have been infected by passing outside buildings, where the source of the causative *Legionella* strain has been found (24). However, some authors suggest that aspiration is more important than inhalation as mode of transmission (25). This is likely in a hospital environment with immunocompromised patients who have poor respiratory tract reflexes. In such cases, drinking water contaminated with *Legionella* is a possible risk. In the Philadelphia outbreak, only drinking water at the hotel was the statistically significant relationship that could be associated with disease in a case control study (4). However, Inhalation of aerosols is perhaps an important factor (24). The sources of transmission are cooling towers, evaporative condensers, whirlpool spas and showers. An air conditioning system is only hazardous if a cooling tower or evaporative condenser is positioned in such a way that the generated aerosol can pass into the air intake of a building or be directly transmitted to a passerby bystander (24).

Inhalation or microaspiration of *amoeba* could be a potential risk, since one single *amoeba* might harbor more than 1000 *Legionella* cells (12). Moreover, intracellular growth in *Acanthamoeba castellanii* affects monocyte entry mechanisms and enhances the virulence of *L. pneumophila* (21, 22). Thus, it is possible that infection in humans may require the presence of both *Legionellae* and an amoebal host (22). This might explain why the attack rate in outbreaks of Legionnaires' disease (LD) is low, despite the

presence of *Legionella* spp in the plumbing system. Breiman *et al* described the occurrence of LD cases related to the presence of both protozoa and *L. pneumophila* (26).

Before November 2014, there were no cases of LD transmitted from person to person reported, indicating that *L. pneumophila* is not completely adapted to infect humans (27). In November 2014, the first case was reported for person to person infection of LD in Portugal (28). This case for mother taking care of her son having LD who was maintenance worker at an industrial cooling tower. Both mother and her adult son died from severe complications of LD. The mother had never travelled to the site of the outbreak. Also, genetic analysis showed that the mother was infected by the identical strain of her son (28).

1.2.3 Molecular identification and genomics of *Legionella*

Molecular identification of *Legionella* spp is possible with an array of fast and widely used techniques. Several methods based on the analysis of DNA are available to classify *Legionella* at the genus- and species-level such as PCR amplified DNA profiles (29) Restriction Fragment Length Polymorphism (RFLP) PCR (30, 31) and sequencing of the mip gene (32). Detection and identification techniques aim at the clinical diagnosis of legionellosis and identification of health risks in freshwater systems. However, until now the PCR assays and specific sequence techniques have good discriminative power (32).

16S rRNA sequencing is the most common and widely used technique for phylogenetic classification of *Legionella* spp (31). Sequencing of the 16S rRNA gene for *Legionella* revealed high taxonomic resolution. Because *Legionella* exhibits multiple ribosomal operon copies with very little sequence heterogeneity of the most widely used fully sequenced genomes of *L. pneumophila* reference strains, i. e. Philadelphia, Paris and Corby (31).

In fact, the discrimination of 16S rRNA gene on the species level is not good enough as on the genus level. Thus, phylogenetic studies using *mip* gene (33, 34), *rpoB* gene (35) and the hypervariable 23S-5S ribosomal intergenic spacer region (36) have been used among *Legionella* spp. In comparison to 16S rRNA sequencing, the mip gene have been widely used for *Legionella* spp revealing that it has a higher variability (32). Moreover, there is mip gene sequence database (http://www.hpa-bioinfotools.org.uk/mip_ID.html) to compare any unknown sequence with all available species. However, *rpoB* allows distinguishing the subspecies *fraseri* and *pneumophila* which is impossible by *mip* and 16S rRNA sequencing. The same discriminatory power has been observed for 23S-5S region allows differentiating the three subspecies *pneumophila*, *pascuelli* and *fraseri*. The progressive availability of genomic information for *Legionella* with allow the identification of a combination of genes with wide discriminatory power for developing an appropriate Multi-Locus Variable number of tandem repeat Analysis(MLVA).

Identifying *L. pneumophila* on a strain level helps to identify environmental sources giving arise to cases of legionellosis. Also, knowing the population structure of *L. pneumophila*, studying genetic diversity and clonal expansion helps in long-term epidemiological analyses of microbial populations. A lot of molecular techniques were developed for molecular typing and identification of *L. pneumophila* on strain level, such as Amplified Fragment Length Polymorphism (AFLP), Restriction Fragment Length Polymorphism (RFLP), Sequence Based Typing (SBT) and MLVA (37, 38). The last two techniques are widely used nowadays and were developed by members of The European Working Group for *Legionella* Infection (EWGLI) and Pourcel *et al* 2007 (38), respectively. The gold standard for high resolution genotyping of *L. pneumophila* isolates is an SBT approach using 6 genes (*flaA, pilE, asd, mip, mompS* and *proA*).

MLVA has exhibited higher discriminatory power than SBT (38, 39). Also, rapidity, typeability, reproducibility are better and a large database is available (http: //bacterial-genotyping.igmors.u-psud.fr/Legionella2006/help.htm) on the web. This makes MLVA an excellent choice for studying and typing *L. pneumophila*. Molecular typing, either SBT or MLVA of *L. pneumophila* is not good enough to know 100% the source of outbreak. Phylogenetic top-down tree shows the detection limit for each molecular analysis techniques. Deep molecular techniques, like whole genome sequencing (WGS), are needed to analysis *L. pneumophila* on a clonal level **(Figure.3)**.

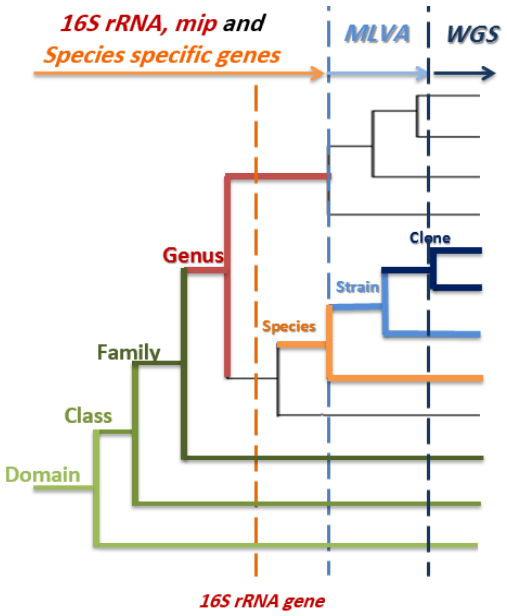


Figure 3. Upper limits of taxonomic identification based on DNA from the envrionment or cultured isolates. For High resolution *L. pneumophila* genotyping sequncing of spefic genes (e.g. mip), MLV and WGA are the most appropiate diagnostics tools. MLVA = Multi Locus of Variable-number Tandem Repeats Analysis WGA = Whole Genome Sequencing . Adopted from Höfle *et al*. 2017

The *L. pneumophila* genome size is 3.3-3.5 Mbp on a single circular chromosome (40). Also, *L. pneumophila* has more 3000 of genes. This reflecting the capacity of *L. pneumophila* to adapt different environments and hosts. High homogeneity of *L. pneumophila* with 38% CG content of coding regions and average length of coding

sequencing. *L. pneumophila* str. *Paris* and *Lens* contains plasmid and *L. pneumophila* in general can be excised on contained plasmid (40).

1.3 Legionellosis and country-specific epidemiology for the West Bank

The term legionellosis or Legionnairs' disease (LD)cause by two forms of disease: 1) a flu-like illness called Pontiac fever named after the city Pontiac, Michigan, USA where an explosive outbreak had occurred in this city (41) and 2) a sever fatal pneumonia with multisystem failure (5). Diabetic, cancer, AIDS and renal failure patients are at risk for LD. Also, elderly people, smokers, males and immunocompromised people or other risk groups (42). More than 650 outbreaks were reported until 2005 worldwide. In the last decade, the mortality rate is decreasing due to earlier detection and better treatment (43).

According to the current epidemiological data available from the world, *L. pneumophila* has 15 serogroups, but majority of LD is caused by sg.1. Around 84% worldwide and 87% in Europe has reported cases of *L. pneumophila* sg.1 LD (44, 45). The only exceptions are Australia and New Zealand having a lower percentage (50%) of *L. pneumophila* and exclusively 30% of LD caused by *L. longbeachae* (44). In Europe, 6,943 cases of LD were reported during 2015 (45). France, Germany, Italy, Portugal and Spain accounted 74% of all notified cases(45). Whereas in the Middle East and in Arabic countries there is a shortage of epidemiological data for *Legionella* spp.

In Israel, limited epidemiological data are available (46, 47). Two recent studies described the molecular epidemiology of *L. pneumophila* by SBT (48) and MLVA (49). Moran-Gilad *et al* (48) showed that 71.4% and 21.7% of the clinical and environmental strains were sg.1 respectively. However, sg.3 was responsible for 14.2% of the infections in Israel. The common sg.3 reported in drinking water as described by Yarom *et al* 2010 (50). Rodriguez-Martinez *et al*. 2015 (49) described the clonal populations of a set of environmental isolates of *L. pneumophila* from freshwater supply system of the Oranim campus in Haifa. Also, they found Sg.1 and Sg.3 dominated in this DWSS (49).

In Turkey, *Legionella* spp contaminated 13.3% of hotels water system in Alanya. Sg.6 was the dominant serogroup (55.6%) followed by sg.1 (22.2%) from the environmental samples (51). In Northern Africa, *Legionella* spp was identified and genotyped by PFGE from Tunisian hot water springs (52). Also, environmental *L. pneumophila* isolates were typed by PFGE from seven Moroccan towns (53).

In Kuwait, *Legionella* was detected and quantified using RT-PCR from water systems in residential facilities. Serological typing of *L. pneumophila* isolates is in accordance with study of Yarom *et al* 2010 (50) and Rodriguez-Martinez *et al* 2015 (49) that sg.3 is the dominant serogroup in Kuwait (54). Recently, studies by Al-Matawah *et al* 2012 & 2015 reported the presence of *L. pneumophila* in air conditioned buildings in Kuwait by using molecular techniques (55). More recently, the European Centre for Disease Prevention and Control (ECDC) reported increasing in LD in EU travelers returning from Dubai, UAE. 26 cases of LD have been reported between October-December 2016 (56).

No study has been done previously in the West Bank employing the molecular identification, serotyping and genotyping of *Legionella* spp in freshwater samples which are considered as opportunistic pathogens causing Legionellosis. Only one study by Shareef and Mimi 2008 studied hospitals tap water systems but they only used culture-based microbiological techniques for isolation and identification of *L. pneumophila* (57).

In Palestine, there is no legislation, rules and protocols from the PWA about *Legionella* spp. Generally, they did not identify and test *Legionella* in freshwater system. Moreover, the Ministry of Health (MOH) does not have any legislation, rules, protocols and procedures how to identify legionellosis. In fact, urine antibody-based identification is used and rapid serological tests (IgG and IgM) are available in the main medical laboratories but these tests are rarely ordered by physicians.

The traditional protocol for patient suspected pneumonia is chest X-ray and Complete Blood Count (CBC). Fluffy infiltrates in the chest by X-ray and granulocytosis in CBC indicate an atypical pneumonia with symptoms like malaise, cough, chills, dyspnea, headache and chest pain (58). Mainly, patients are then treated with appropriate antibiotics such as fluoroquinolones and macrolides (59).

1.4 The genus *Legionella*, its ecophysiology and molecular taxonomy

Legionella is a genus comprising about 60 species mostly of aquatic origin and with a large fraction of pathogenic species (60). Most relevant for human health and man-made freshwater systems is *L. pneumophila* that is the most relevant causative agent of an atypical pneumonia, Legionnaires ' disease (LD), and Pontiac fever, a self-limiting flu-like disease (61). Anthropogenic fresh water systems are considered as the major source for *Legionella* infections (62, 63). Co-infections with aquatic bacteria of LD patients hint on a co-transfer of bacteria from freshwater to the patient presumably via protozoa and/or their bacteria containing vesicles (64, 65). LD is transmitted by inhalation or aspiration of aerosols from contaminated environmental sources. Thus, when a case of LD is diagnosed other persons in the same environmental conditions might be at high risk to become infected from the same environmental source. The source of such an outbreak can be determined by molecular characterization of *Legionella* isolates from the patient and from environmental samples (37).

Environmentally, *Legionellae* are found in soil and water in a symbiotic close association with amoebae (14, 66). *L. pneumophila* life cycle alternates between a replicative intracellular phase and a transmissive extracellular phase in response to different life conditions (67). When the conditions are favorable, it can replicate within the host cell vacuole. Contrary, when the conditions are limited, it becomes cytotoxic, resistant to osmotic stress, sensitive to sodium and competent to evade phagosome-lysosome fusion (67). These traits enable *L. pneumophila* to survive and replicate in the environment.

In humans, the inhalation of *L. pneumophila* contaminated aerosols may cause LD (4). It replicates inside alveolar macrophages and epithelial cells (68, 69) in a similar

mechanism to its invasion of amoeba host (68). After phagocytosis by mammalian and amoeba cells, *L. pneumophila* modulate the biogenesis of their vacuole which becomes surrounded by mitochondria and rough endoplasmic reticulum (70) then, *L. pneumophila* starts replicating.

Many factors promote *L. pneumophila* virulence in the alveolar macrophages, survival and replication in amoeba. The most important one is Dot (Defect in Organelle Trafficking)/icm (intracellular multiplication). Supported by the Dot/Icm type IV secretion system, the bacteria avoid phagolysosome fusion and replicate within alveolar macrophages and epithelial cells (71, 72). The Dot/Icm is essential for pore-formation mediated lysis of the host cell (69, 73).

In vitro, *L. pneumophila* in Post Exponential (PE) phase is most infectious and more resistant to many stresses. *L. pneumophila* in PE phase displays shortened cell body, flagellin expression, pigment accumulation and reduced sodium sensitivity. These traits, together with virulence markers such as cytotoxicity, intracellular growth and phagocytosis are recognized as the transmission traits of *L. pneumophila* (74, 75).

The West Bank is a semi-arid region in the Middle East with hot and dry summers and cool winters with substantial water scarcity problems. Main precipitation falls in winter leading to an often only partial recharge of groundwater aquifers. The source for drinking water is mostly groundwater that is pumped into a storage reservoir and chlorinated, before delivered to the drinking water distribution system (DWDS). Due to frequent water shortage and supply interruption, water is stored in private containers, mostly on the roof, by the end users (3). All these factors, but especially the warm climate, intermittent water supply and storage in roof containers may cause hygienic water problems in general and may lead to high *Legionella* abundance in the drinking water as used by the end user.

The contamination of hospitals' water systems with *Legionella* is considered as a high risk for patients especially for those with severe diseases hospitalized for a long period of time. It is well known that LD is an important cause of hospital-acquired pneumonia (76). Presence of *Legionella* in water distribution systems actually is a serious health risk to hospital staff and patients, but the magnitude of the problem is generally unrecognized (24, 76, 77). As for Palestine the problem is that clinical awareness about the prevalence of *L. pneumophila* or LD is lacking. Furthermore, there are no specific guidelines for *Legionella spp* surveillance and thus protection from exposure in hospitals or public buildings.

Cultivation Dependent Analysis (CDA) (culturing) is the standard and recommended technique used for environmental surveillance of *L.pneumophila* (78). One major hassle in the isolation and quantification of *L. pneumophila* by culture is the Viable But Non Culturable (VBNC) state and its overgrowth by competing bacteria (61, 79, 80). On the other hand Cultivation Independent Analysis (CIA) are rapid, sensitive and widely used methods for detection and identification of *L. pneumophila* (81). Multi Locus Variable number of tandem repeats Analysis (MLVA) is the molecular typing method used to

assess the genetic diversity among *L. pneumophila* isolates and identification of *L. pneumophila* at the strain level. MLVA genotyping has exhibited high resolution in comparison with other typing methodologies for multiple microorganisms (82-84). In the case of *L. pneumophila,* MLVA genotyping presented higher discrimination than Sequence Based Typing (SBT) (38, 39) and a very high concordance with the gold standard method as described by Pourcel *et al*. (2007), Visca *et al*. (2011) (38, 39).

The standard method for genotype assessment of *L. pneumophila* is sequence based typing (SBT) (60). Sobral *et al*. and Visca *et al* (2011) (39, 85) showed that multi-locus variable number of tandem repeat analysis (MLVA), a less laborious method, can be very well matched with the sequence types (ST) generated by SBT. In addition, the genotypes of ST1 are better resolved by MLVA. ST 1 resolution had been shown to be of special relevance for drinking water in the study by Rodriguez *et al*. (2015) (49), for environmental studies as well as for virulence aspects (49, 86). Furthermore, ST1 is considered of high relevance for man-made freshwater systems and human health on a global scale (87, 88).

LD is primarily caused by the species *L. pneumophila* that is widely distributed in the natural and man-made freshwater environments. Differentiation of isolates from this species by molecular typing methods is important to identify the source of outbreak. Many molecular typing methods have been used for epidemiological purposes, including the commonly used method Sequence Based Typing (SBT) (60) and Multi-Locus of Variable Number of Tandem Repeats Analysis (MLVA) with slightly higher discriminatory power (38). However, the discriminatory power of MLVA is still limited by their utility for investigating the intra-clonal level of diversity among outbreaks and often cannot distinguish between outbreak and non-outbreak isolates (39, 85). Nowadays, whole genome sequencing (WGS) has been successfully applied for *L. pneumophila,* which provides the highest conceivable discriminatory power to distinguish outbreak from non-outbreak isolates (89, 90).

L. pneumophila strains Philadelphia, Paris and Lens were the first whole genomes that have been sequenced (40, 91). These *L. pneumophila* strains contain one circular chromosome of a size 3.3 to 3.5 Mb and have more than 3000 protein coding genes representing 88% of the genome sequence. In fact, the genome of *L. pneumophila* has been shown to be highly plastic (40), including Horizontal Gene Transfer (HGT), mobile elements and Genomic Islands (GIs) (92). Generally, most WGS studies focused on outbreak isolates (93, 94) if they are not dealing with the genetics and evolution of *L. pneumophila* (87). The use of WGS for isolates typing is generally based on either Single Nucleotide Polymorphism (SNPs) comparison or gene by gene comparison (95, 96). In this study, we used MLVA to assess the genotypic characteristics and biogeographical distribution of *L. pneumophila* strains isolated from freshwater system in hospitals across the West Bank and revealed four clonal complexes, some of them were also prevalent in Germany (97).

Ecologically, the abundance of *Legionella* is considered to be enhanced by high water temperature and low chlorine concentrations (98-100). However, recent studies have

indicated that it is of high value to study the ecology of *Legionella* on the genotype level. Drinking water distribution systems can be dominated by a few or even a single genotype (49, 85). In addition, the study by Rodriguez *et al* (2015) performed in a similar climatic region (around Haifa, Israel) hints on a link between the genotypic composition of *L. pneumophila* and the abundance of *Legionella* species. In their study a specific genotype (MLVA-Gt4) closely related to *L. pneumophila* strain Paris showed a correlation with very high *Legionella* plate counts and surprisingly low water temperature (around 22°C). The authors suggested that specific genotypes might act as triggers of *Legionella* abundance, even at low temperature. In addition, genotype assessment is also of interest for tracking the source of *Legionella* infections and the virulence itself is considered to be genotype dependent (86, 101).

1.5 Objectives of the Thesis and used approaches

The overall goal of the Thesis to advance current knowledge on the environmental reservoirs of *L. pneumophila* in fresh water systems, the environmental, geographical, physiochemical, virulence and genomic factors affecting their clone specific occurrence and virulence to humans. The following three detailed objectives are pursued in this Thesis:

Objective 1: Assessment of the biogeographical distribution of *L. pneumophila* isolates according to genotype compostion across the West Bank and identification of its environmental drivers.

Objective 2: Determination of virulence potential of *L. pneumophila* isolates from hospital water systems according to genotypes and clonal complex classification.

Objective 3: Genome comparison of *L. pneumophila* isolates from the West Bank and Germany regarding different criteria: identification of virulence genes and genomic islands.

In the current study these detailed objectives were approached by a seasonal assessment of *Legionella* abundance in water and biofilm in eight sampling sites covering the whole West Bank was approached. The eight sampling sites were eight hospitals in the West Bank, i.e. tap water and the respective biofilms were sampled. The abundance of *L. pneumophila* was assessed by cultivation and PCR concomitantly with a record of major bacteriological and physico-chemical parameters of the drinking water during a period of 2.3 years. *L. pneumophila* isolates obtained from water and biofilm were identified by 16S rRNA partial sequencing. Genotypes of *L. pneumophila* were assessed by a high resolution MLVA using 13 loci (39). Correlation analysis and principal component analysis (PCA) were used to identify the niches of relevant *L. pneumophila* genotypes and to identify environmental drivers of *L. pneumophila* abundance in water and biofilm. The study promotes a genotype-based ecology of *L. pneumophila* and sheds light on so far not yet considered mechanisms of *L. pneumophila* prevalence at the level of individual genotypes. Furthermore, the virulence of the *L. pneumophila* genotypes and VACCs was assessed and compared by five different in-vitro assays.

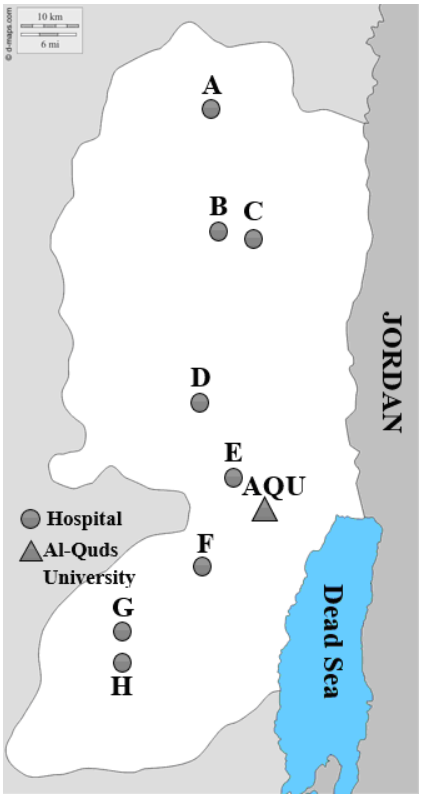
Lastly, the study aimed at comparing representative *L. pneumophila* isolates from two very different geographical areas namely the West Bank and Germany at the level of their whole genome sequences to reveal their differences across large and small geographical distances, their genome dynamics and to understand their genetic make-up for virulence. For the achievement of this aim genome sequences from *L. pneumophila* reference strains and our isolates were compared using SNPs comparison. This analysis is based on complete genome sequencing of already published reference genome strains of *L. pneumophila* strain Lpn-Paris (40), Lpn-Corby (102), Lpn-Philadelphia (91), Lpn-Alcoy 2300/99 (103), Lpn-Lorraine, Lpn-Thunderbay (104), Lpn-D7630, Lpn-Pontiac (105), Lpn-OLDA (105), Lpn-ATCC43290 (106) and two isolates. In addition to deep sequencing on the Illumina platform, an environmental isolate from the West Bank and a clinical isolate from Germany were sequenced using the Pacific Bioscience (PacBio) platform. Finally, the genome dynamics of *L. pneumophila* strains was investigated by analyzing Genomic islands, mobile elements and virulent genes of the whole genome sequences.

2. Materials and Methods

2.1 Study sites, water and biofilm sampling

Water samples and biofilm swabs were sampled six times during the period from October 2012 till December 2014 from eight hospitals across the West Bank (WB). The six samplings covered twice the main seasons, i.e. spring (March-May), summer (June-August), and autumn (October - December).

Sampling was achieved on tap water and biofilm of faucets and shower heads of eight different hospitals **(Figure 4):** Hospital A (coordinates: 32° 27' N, 35° 17' E), hospital B (32° 13' N, 35° 14' E) and hospital C (32° 13' N, 35° 15' E) from Northern WB, hospital D (31° 53' N, 35° 12' E) and hospital E (31° 46' N, 35° 14' E) from Central WB, and hospital F (31° 42' N, 35° 11' E), hospital G (31° 33' N, 35° 4' E) and hospital H (31° 31' N, 35° 5' E) from Southern WB. Also, samples were taken occasionally from Al-Quds University (AQU) main campus, Abu Dies, Jerusalem (31° 45' 18.07'' N, 35° 15' 37.614'' E). These samples from Al-Quds were not included in the overall comparison on *Legionella* ecology but only used for comparison.



Cold and hot water (if available) was collected from the main water source of each site and a defined set of biofilm swabs from faucets, showerheads and hoses. The sampled drinking water is considered as representative for the cities of Jenin, Nablus, Ramallah, Jerusalem, Bethlehem and Hebron, going from the north towards the south across the West Bank.

Figure 4: Sampling map of the eight hospitals and Al-Quds University in the West Bank, Palestine.

2.2 Physico-chemical analysis of the water

Cold and hot water samples were tested for temperature, pH and conductivity using probes and chlorine (Quantofix) directly upon collection. After return to the laboratory water samples were tested for total iron, nitrate, nitrite, ammonia, copper, phosphate, zinc, carbonate hardness, total hardness using Quantofix sticks (Macherey-Nagel GmbH & co.KG, Germany). Magnesium and calcium concentrations were measured photometrically using Nanocolor assays (Macherey-Nagel GmbH & co.KG, Germany). Data on turbidity, bicarbonate, chloride, sulfate, total dissolved solids (TDS) and fluoride were provided by the Palestinian Water Authority (3).

2.3 Cultivation dependent analysis of water and biofilm

A total of 72 water samples were collected from the main hospital tanks in sterile 1L plastic bottles after a brief flow time (2-3 min), to permit clearing of the service line. One liter of each cold and hot water was collected for Heterotrophic Plate Count (HPC) and one liter of each cold and hot water was collected for *Legionella* count from the hospitals. To neutralize residual free chlorine, 0.5ml of 0.1N sodium thiosulphate was added in the sterile bottles for *Legionella* count analysis (107).

Biofilm swabs were sampled from the anterior surfaces of faucets, showerheads or shower hoses in all hospitals wards. Per sampling and sampling site 20 biofilm swabs were taken, except for the first sampling when 44 swabs were sampled per site in order to check the variability per sampling site. Swabs for *Legionella* identification were processed immediately by culturing on GVPC (M809, Himedia, India) medium and procedure based on ISO 11731:2004 (108).

From all water and biofilm samples with visible growth on *Legionella*-like colonies on agar plates, representative isolates were chosen and purified. Isolates were later characterized by *L. pneumophila* specific PCR, 16S rRNA gene sequence, serogroup and genotype assignment using MLVA (see below). A subset of MLVA genotypes was characterized by sequence based typing.

2.4 Cultivation independent analysis of water and biofilm

Five liters of each cold and hot water were collected per sampling and site from the main water source for DNA extraction. Water samples were filtered onto sandwich membrane filters composed of nucleopore-filter (Nuclepore Track-Etch Membrane, MB 90mm, 0.2µm, Whatman, England) and glass fiber-microfilter (GF/F) (GFF, 90mm, Whatman, England). Filter was stored frozen for later DNA extraction.

For biofilm, 5 swabs were taken per sampling and site from the anterior surfaces of faucets, showerheads or shower hoses using sterile cotton swabs (Cotton Tipped Applicator, China). Swabs were stored frozen for later DNA extraction.

For the extraction of DNA from the filter sandwiches and the swabs, a modified DNeasy protocol (Qiagen 69506, Germany) was used. Briefly, sandwich filters were cut into small pieces and incubated with enzymatic lysis buffer (20 mM Tris-HCl, 2 mM EDTA, 1.2% Triton X-100 (pH 8.0)) containing 10 mg/ml lysozyme for 60 min in a 37°C water bath. After the addition of AL buffer from the kit, the samples were incubated at 78°C in a shaking water bath for 20 min. After filtration through a cell strainer 100 µm (DB falcon 352360, USA), absolute ethanol was added to the filtrate (ratio of filtrate to ethanol is (2:1)), and the mixture was applied onto the spin column of the kit. After this step, the protocol was followed according to the manufacturer's instructions.

Three PCRs with different targets were carried out as follows; for the detection of any bacteria the complete 16S rRNA gene primers, Com1F (5'-CAGCAGCCGCGGTAATAC-3') and Com2R (5'-CCGTCAATTCCTTTGAGTTT-3') were used. For the classification of *Legionella* like colonies a *Legionella* genus-specific PCR using primers Lgsp17F (5'-GGCCTACCAAGGCGACGATCG-3') and Lgsp28R (5'-CACCGGAAATTCCACTACCCTCTC-3') and an *L. pneumophila*-specific PCR using primers Lp-16S_246-248F (CCTGGGCTTAACCTGGGAC) and Lp-16S_246-248R (CTTAGAGTCCCCACCATCACAT) were applied (109). Each reaction was carried out by using 3µl (100mg/µl) of DNA template in a final volume of 25µl. Amplification was achieved by using PCR-ready Master Mix (GoTaq, Green Master Mix, Promega, USA).

On all samples, used for PCR based analysis, cultivation dependent analysis for *Legionella* was additionally performed to allow a direct comparison. Sequencing of the 16S rRNA gene of the six isolates was confirmed as *L. pneumophila* (≥99.8% 16S rRNA gene similarities). 16S rRNA sequences of *L. pneumophila* isolates were submitted to the GeneBank database (KX778102 - KX778107).

2.5 Genotyping of *L. pneumophila* isolates

For molecular typing of *L. pneumophila* at the strain level. MLVA-8(12) analysis was performed for 180 isolates. DNA extraction was done either directly from living biomass using (Qiagen kit lot No: 69504) according to the manufacturer protocol or from biomass on FTA cards (Whatman, Sigma-Aldrich, Germany). DNA was finally quantified by Nanodrop spectrophotometer (NanoDrop, Thermo Scientific, Germany).

For DNA extraction out of the FTA cards, the area of the card containing the biomass was punched with a puncher into 3 mm circular punches. Punches were transferred to 0.5 ml sterile water (Roth, Germany). It was incubated for 3 min at room temperature and vortexed three times (after water addition, after 1 min and after 3 min incubation). The FTA punch was removed and 1x Tris-EDTA buffer (Sigma-Aldrich, Germany) was added to the water to preserve the DNA from degradation. MLVA-8 and MLVA-12 molecular genotyping assays by capillary electrophoresis were carried out for all isolates as detailed by (38, 39, 97). For the final MLVA thirteen loci were used, i.e. the twelve loci of MLVA-12 (85) plus the one additional locus of MLVA-8 not used in MLVA-12.

2.6 Serogrouping of *Legionella* isolates

The serogrouping of subset of 180 *L. pneumophila* isolates was identified by an agglutination test using (Oxoid DR0800) *Legionella* Latex *L.pneumophila* was serogrouped as Sg1 and Sg 2-14. Moreover, 47 isolates were sent to the National Reference laboratory for *Legionella* infections in Dresden for analyzes by monoclonal antibody subgrouping (110).

2.7 *Legionella* species, *Acanthamoeba castellanii* cells and Cell culture

Sixty *L. pneumophila* environmental isolates from the West Bank-Palestine were used for different virulence assessment tests for the genotypes Gt4(17), Gt6(18), Gt10(93), Gt10(141), Gt12(84), Gt40(47), Gt13(72), Gt9(92) and Gt64(74) and the four clonal complexes (VACC1, VACC2, VACC5 and VACC11).

The virulent clinical isolate of *L. pneumophila* strain AA100 and its isogenic dotA mutant was used in this study and have been described previously (111). *L. pneumophila* str. Corby (CP000675), *L. pneumophila* str. Paris (CIP 107629), *L. pneumophila* str. Philadelphia-1 (ATCC 33152) were used as *L. pneumophila* reference strains.

L. pneumophila strains were grown in Yeast Extract Broth (YEB) until the Post Exponential (PE) phase. All *L. pneumophila* strains were plated on BCYE agar plates and incubated at 37°C for 72 h. To standardize the inoculum for all tests, all strains were additionally pre-cultured in broth until the bacteria had reached the PE phase. For that, 50 ml of 1:2 diluted YEB liquid medium [10 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES), 10 g yeast extract, 0.4g L-cysteine and 0.25g ferric nitrate and 1 g α-ketoglutaric acid per liter of sterile water] were inoculated with fresh biomass and incubated at 37°C with agitation (100 RPM) for 48 h. After incubation, pre-cultures were diluted with fresh medium and the optical density at 600 nm (OD 600nm) of all the strains were set to 0.01 (equivalent to cell density of 10^7 cells/ml) to standardize the initial inoculum. Each strain was studied using triplicates for each assay. THP-1 macrophages were cultured in RPMI 1640 medium containing 10% bovine fetal serum (Gibco). THP-1 macrophages were differentiated with phorbol 12-myristate 13-acetate (PMA) (sigma) for 24 h before use as described previously (97, 111). *A. castellanii* cells were maintained in Peptone Yeast Extract Glucose (PGYE) medium.

2.8 Pore-Forming Mediated Cytotoxicity assay

The virulence of the previously mentioned isolates were assessed by contact-dependent pore formation of sRBC's by *L. pneumophila* at multiplicity of infection (MOI) of 25 after 2 hours of bacterial-sRBC's contact, as described previously (69, 112). The release of hemoglobin from lysed red blood cells was measured by spectrophotometry at 415 nm. Pore forming cytotoxicity was expressed as percentage of hemolysis compared to 100% fully hemolyzed blood cells.

L. pneumophila isolates were plated and cultured on BCYE medium for 72 h. Packed sRBCs (Catalog No.: IC100-0210, Innovative research, USA) were prepared for the assay. SRBCs (250 ul) were diluted in 50 ml of Phosphate Buffered Saline (PBS). The cells were then washed three times by centrifugation for 15 min at 2,000 RPM at 4°C, 3 times. The sRBCs were re-suspended in 25ml RPMI-1640 modified medium (Gibco, Germany) and counted using a Neubauer microscopy chamber. The number of SRBCs was adjusted to a concentration of $1x10^8$ and $1.5x10^8$ cells/ml. *L. pneumophila* isolates were suspended in 7 ml PBS, and then diluted to 1:20 and the O.D was measured at 600nm using spectrophotometer Nanocolor Vis (Macherey-Nagel, Germany). *Legionella* cultures OD 600 nm was adjusted to MOI of 25. Bacterial cells were pelleted at 13,000 RPM for 2 min. Then, 0.5 ml of the sRBCs ($1x10^8$ cells/ml) were added to the bacterial cells, Bacterial-sRBCs mixture were centrifuged at 13,000 RPM for 2 minutes and the mixture incubated at 37°C for 2 h. After incubation, the mixture was re-suspended thoroughly and OD was measured at 415 nm in order to determine the amount of remaining intact RBCs.

2.9 *Acanthamoeba castellanii* infection assay

The virulence of the previously mentioned isolates was assessed by infecting *A. castellanii* by *L. pneumophila* at MOI of 10 for 24hrs as previously described (113). The percentage of survived *A. castellanii* was calculated as (*A. castellanii* infected with *L. pneumophila*/ *A. castellanii* concentration of positive control well) X 100%. Then, the percentage of killed *A. castellanii* was measured as (100% - the percentage of survived *A. castellanii*).

A. castellanii was cultured in PYGE medium at 30°C for 72 h in cell culture flask. *A. castellanii* cells were washed three times with PYGE medium and 5 ml of PYGE medium were added for the detachment of *A. castellanii* cells from the flask surface. Then, *A. castellanii* cells were counted in Neubauer microscopy chamber and cell concentrations were adjusted to $1.25x10^4$ cell/ml suspension in PYGE. Two ml of *A. castellanii* cells were pipetted in each well of twelve well plates. Then, the plate was centrifuged for 2min at 500xg. The plate was incubated for 3 hours at 30°C in order for cells to adhere properly to the wells surface. Each *L. pneumophila* isolate was inoculated in 20ml YEB with initial OD600 of 0.1. Bacterial cultures were incubated at 37°C until they reach the post exponential phase of growth and then Centrifuged at 2500xg for 15 min. Supernatant was discarded and bacterial pellet was washed with PBS and re-suspended in PYGE medium to a final OD600 of 0.078 (equals to $6.25x10^7$ CFU/ml). The appropriate volume, which corresponds to a 1:10 ratio with *A. castellanii,* was calculated and added to the 12 well plates. Plates were centrifuged for 5 min at 500xg to allow contact between *A. castellanii* and *L. pneumophila*. Each *L. pneumophila* isolate was added in triplicate wells containing identical *A. castellanii* concentrations. Positive control (only *A. castellanii* in PYGE medium) was added in triplicate wells. Negative control (only PYGE medium) was added in two wells. Then, plates were incubated at 30°C for 24h. Afterwards plates were placed on ice for 30 minutes for proper detachment of *A.*

castellanii cells from the bottom of the wells. Intact *A. castellanii* cells were counted in Neubauer chambers.

2.10 Heat shock and sodium chloride tolerance assays

The virulence of the previously mentioned isolates were assessed by heat shock at 60°C by (1.00E+08) *L. pneumophila* cells/ml after 30 min incubation as described previously (74). In detailed, *L. pneumophila* isolates were plated and cultured on BCYE medium for 72 h. *L. pneumophila* isolates were suspended in 5 ml PBS, and then diluted to 1:100 and the O.D was measured at 600nm using spectrophotometer (Human, Germany). *L. pneumophila* cultures O.D600 were adjusted to 0.1 equivalents to (1.00E+08) cells. Then, *L. pneumophila* cells suspension were incubated at 60°C for 30 min. Then, *L. pneumophila* cells were cultivated on BCYE medium for 72 h. Heat resistance of *L. pneumophila* was calculated as [(stressed sample CFU/ml)/(control sample CFU/ml)] × 100. Also, the virulence assessment of (1.00E+08) *L. pneumophila* cells/ml with 100mM NaCl was performed as previously described (75, 114). *L. pneumophila* isolates were plated and cultured on BCYE medium for 72 h. *L. pneumophila* isolates were suspended in 5 ml PBS, and then diluted to 1:100 and the O.D was measured at 600nm using spectrophotometer (Human, Germany). *L. pneumophila* cultures O.D600 were adjusted to 0.1 equivalents to (1.00E+08) cells. Then, *L. pneumophila* cells were cultivated on BCYE containing 100mM NaCl medium and BCYE medium as control 72 h for CFU counts. Sodium sensitivity was calculated as [(BCYE-100mM NaCl CFU/ml)/(BCYE CFU/ml)] × 100.

2.11 Cytopathogenicity of THP-1 Macrophages assay

The virulence of the previously mentioned isolates were assessed by infecting THP-1 macrophages by *L. pneumophila* at MOI of 10 after 24 h of *L. pneumophila*-THP-1 macrophages infection as previously described (68, 97). The relative degree of cytopathogenicity was expressed as percent of inhibition compared to non-infected cells; calculated as (Y= [(K-Y)/K]*100. As K: mean OD of non-infected cells and Y: OD of infected cells.

Bacterial density was assessed by the absorbance at 600 nm with a spectrophotometer Nanocolor Vis (Macherey-Nagel, Germany). Briefly, 10 ml of bacterial culture was centrifuged at 3500 RPM for 15 min, the pellet was suspended in 5 ml of PBS and the optical density at 600nm was adjusted to 0.7 equivalents to 10^9 bacteria/ml. The number of cells was checked by plating on BCYE agar plates and counting the CFU. Bacteria was diluted with RPMI-1640 modified medium (Gibco, Germany) to the desired MOI. THP-1 cells (ATCC TIB-202TM) were maintained in RPMI-1640 modified medium (Gibco, Germany) containing 10% fetal bovine serum (FBS) (Gibco, Germany) and 2mM L-Glutamine (Gibco, Germany) in a humid atmosphere containing 5% CO2 at 37°C. Before each infection experiment, exponentially growing THP-1 monocytes were washed with complete medium, counted and incubated with phorbol-12-myristate-13-acetate (PMA) (Sigma Aldrich, Germany) at a concentration of 100ng/ml for 48h in 5%

CO_2 at 37°C to induce maturation of the monocytes into macrophage-like adherent cells. Adherent cells were washed three times with PBS prior to infection.

Intracellular multiplication assays of *L. pneumophila* strains were carried out as previously described (68, 111). Briefly, differentiated THP-1 macrophages ($2X10^5$ cells per well) in 24-well plates were infected in triplicate with 1 ml of bacterial suspension at a MOI of 10. After 1 h incubation, cells were washed twice with PBS and incubated with 1 ml of RPMI medium containing 100 ug /ml gentamycin (Sigma-Aldrich, Germany) for 1 h at 5 % CO_2 at 37°C to kill extracellular bacteria. Infected macrophages were lysed after 1 h, 24 h and 48 h using 1 ml of 0.25% Triton 100X (Sigma-Aldrich, Germany) for 20 sec. Dilutions of the lysate were plated in triplicates in BCYE agar plates and incubated at 37°C to determine the CFU. Cytotoxicity assays were carried out by infecting THP-1 macrophages seeded in 96- well plates at a density of 10^5 cells per well with *L. pneumophila* strains at MOI 10 during 1 h at 5% CO_2. To measure the percentage of viable macrophages after the infection, a colorimetric assay based on the oxidation-reduction of resazurin acid (Alamar blue) (Sigma Aldrich, Germany) was used. After infection and washing with PBS, macrophages were incubated with 200 µl of RPMI medium supplemented with 10% Alamar blue (Sigma-Aldrich, Germany). Absorbance was measured at 600 nm after 24 h using a Synergy 2 microplate reader (Biotek, Germany).

2.12 Selection of strains and reference genomes

A total of 180 MLVA-8(12) genotyped environmental isolates from the West Bank and 260 clinical and environmental isolates from Germany were used for this study (97). The selection of a subset of *L. pneumophila* isolates for whole genome sequencing faced many challenges since we had a large set of isolates (440 isolates) analyzed by MLVA-8(12) typing technique.

Different criteria have been discussed before selection. Reasonably, the most important criterion was the clonal relatedness based on MLVA-8(12) genotyping. Using thirteen selected loci on the *L. pneumophila* genome has high discriminatory power (ID=0.964, 95% CI 0.949-0.980) distinguishing *L. pneumophila* strains at the clonal level (85, 97). Thus, this relatively high resolution in identification and classification of *L. pneumophila* makes MLVA genotyping the first rank for selection. Genotype abundance, source of sample and site of isolation are in the second rank. Genome sequences of seventeen *L. pneumophila* reference strains retrieved from GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/index.html) were used as reference sequences. Sixteen environmental isolates from the West Bank were selected. Furthermore, fourteen clinical isolates across Germany and eight environmental isolates from Helmholtz Center for Infection Research (HZI) cooling tower, Braunschweig Germany were used in this study (Table A1). Also, the two *L. pneumophila* isolates (A194_Gt40(47)_Ps and H3_Gt14(31)_D) were carefully selected for PacBio whole genome sequencing for a robust comparison with the newly sequenced genomes.

2.13 Illumina Hiseq and Pacific Bioscience genome sequencing, assembly and annotation

Nowadays, several DNA sequencing platforms are available as e.g. the Illumina Hiseq and PacBio sequencers. The output varies considerable concerning the sequencing method (115). The Illumina Hiseq 2500 is a high-throughput sequencing system. The Illumina Hiseq displays a high sequencing accuracy with error rates smaller than 0.5 percent, but read lengths are limited to a maximum of 2x250 bp with throughput 125-150Gb and 600M reads with $40 cost per Gb (116, 117). In contrast, the PacBio system provides much longer sequence reads of an average of 10 kbp with high throughput up to 7Gbp and around 350,000 reads (117). This 3rd generation sequencing technology is based on Single Molecular Real-Time (SMRT) sequencing. About 150,000 up to one million zero-mode wave guides per SMRT cell are provided on the RSII and Sequel platforms, respectively (118, 119). Meanwhile, there are hundreds of bacterial genome studies performed with PacBio SMRT sequencing including, epigenetic studies, full length mRNA sequencing, targeted sequencing, *de novo* assembly and whole genome sequencing. This ground-breaking technology provides new discoveries in genomic sciences, understanding biology diseases and outbreaks.

Both above-mentioned techniques were applied in this study. Both platforms are well-established and complementary in terms of read length and error types (116, 119). Thirty six clinical and environmental isolates from the West Bank and Germany were sent to the Illumina Hiseq sequencing. Furthermore, the two environmental isolates A194 and H3 from the West Bank and Germany, respectively were selected and 5g weighted cells grown in YEB medium were sent to the German Collection for Microorganisms and Cell Cultures (Leibniz-Institute DSMZ) for PacBio complete genome sequencing.

Genomic DNA libraries for the Illumina sequencer were prepared using the NEBNext Ultra kit according to the manufacturer's instructions (NEB, Ipswitch, MA, USA), and sequenced using 100 bp paired-end runs on an Illumina Hiseq 2500. Processed sequence reads for each isolate were assembled using Velvet version 1.2.10 with k-mer size of 61. Roughly, each *L. pneumophila* genome was targeted to be at 100X coverage within Hiseq sequencing. Hereby, dozens of contigs are retrieved, but no complete genomes.

Genomic DNA of strains A194 and H3 were extracted using Genomic Tip 100/G kit (Qiagen, Hilden, Germany) and a > 4 kb SMRTbell^{TM} template library was prepared according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, CA, USA). Processed sequence reads were assembled into complete genomes using HGAP3 (within SMRT Portal 2.3.0) applying default settings.

Assembled contigs for all sequenced isolates were annotated using the Prokka pipeline v1.11 (120). Prokka relies on external prediction tools to identify the coordinates of CDS, tRNAs, rRNAs, CRISPRs and other genomic features encoded on contigs and chromosomes. All *L. pneumophila* genomes were annotated using a species-specific

database created from the published, manually annotated *L. pneumophila* strain Lpn-Corby (102).

2.14 Genome analysis

The software Parsnp was used for SNP analysis and to construct a phylogenomic tree for the core genome of the completed *L. pneumophila* genomes (A194_Gt40(47)_Ps, H3_Gt14(31)_D and already published *L. pneumophila* genomes) (121). Parsnp generated results were visualized using Gingr (121). Gingr provides an interactive display of multi-alignment variants and phylogenetic trees estimated from the core genome alignment.

Additionally, assembled contigs of *L. pneumophila* isolates along with genome sequences of published strains from GenBank database were aligned, generated and visualized using progressive MAUVE algorithm multiple genome alignment software with default parameters. A comparative analysis among *L. pneumophila* strains has been carried out using MAUVE (122, 123).

Genomic islands (GI) regions were identified using a web-based tool called IslandViewer 4 (http://www.pathogenomics.sfu.ca/islandviewer) (124) that combines the prediction results of three genomic island identification algorithms: Island pick, this is an automated comparative genomics method that selects related genomes for a given query (125). Score Based Identification of Genomic Island Using Hidden Markov Models (SIGI-HMM). SIGI-HMM uses the codon usage frequency table of organisms as their distinctive signatures (126). The SIGI-HMM procedure exploits the difference in codon usage bias between recipient and donor organisms to infer putatively alien DNAs. The visualization tool IslandPath –DIMOB presents a gene map where the potential GI harbored genes, inferred by their significant difference from the GCC content and dinucleotide composition of the genome, and GI specific features such are tRNA genes are indicated by special colors markings. IslandPath-DIMOB identifies GIs through their dinucleotide bias and the presence of mobility genes (127). The outcome of three Genomic Island (GI) prediction programs, i. e. IslandPath-DIMOB, SIGI-HMM, and IslandPick, are integrated by IslandViewer4 into a web interface. It reports better accuracy than any of these three methods alone. Genomic islands present in *L. pneumophila* isolates were predicted by at least one method.

BLASTp was applied to align the amino acid sequences against the Virulence Factors of Pathogenic Bacteria (VFDB) database (128, 129). Amino acid sequences with 75% match identity was chosen and the description of the best hit was assigned as the annotation of predicted gene compared to *L. pneumophila* strain Philadelphia as default bacteria on the webpage.

2.15 Statistical Analysis

Statistical analysis was performed using SPSS 20 and multivariate analyses using PRIMER (version 7.0.7). Non-Normalized data were normalized. Then, repeated measures ANOVA with post-hoc analysis using the Bonferroni test were conducted for determining site-differentiation of water turbidity, HCO3, chloride, SO4, hardness, TDS, magnesium, calcium and calcium/ magnesium ratio and one-way ANOVA was performed to estimate statistical differences among virulence assays and between the isolates from different hospitals. Also, independent t-test was performed to estimate differences among the five virulence assays. All tests were applied at a 95% and 99% level of confidence.

Association between genotypes were calculated using the Similarity Profile Analysis (SIMPROF) (130) based on Spearman rank correlation. If $p<0.05$, groups of samples were considered significantly different. To determine the effect of physiochemical parameters on *L. pneumophila* genotypes, Principal Component Analysis (PCA) were used for visualization of cluster identification. PCA was included 8 parameters. To determine the parameters that affect water quality, PCA was used for visualization of cluster identification. PCA was conducted on conveniences and included 22 variables.

Capillary electrophoresis data analysis and calculation of the number of repeats for each VNTR marker were achieved as described in (38). The numerical code used to designate MLVA-8 and MLVA-12 genotypes, as well as the joint code for MLVA-8(12) genotypes, was continued for the isolates. Null alleles ("0") were assigned when no amplicon was detected. Clustering analysis was performed in Bionumerics (version 5.0, Applied Maths). UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method using a categorical coefficient was applied to define the clusters. MLVA profiles obtained in this study were compared to those from the *Legionella* database and clusters were defined applying as criteria a cut-off of 60% similarity as previously done (Sobral *et al*. 2011). Minimum spanning trees were performed using the categorical coefficient. Simpson's Index of Diversity coefficient was calculated using the online tool provided in http://darwin.phyloviz.net/ComparingPartitions/. To measure the variation of the number of repeats at each VNTR locus, the Hunter-Gaston Discrimination Index (HGDI), which is a modification of the Simpson's Index of Diversity, was calculated using http://www.hpa-bioinfotools.org.uk/cgi-bin/DICI/DICI.pl. Prism 5 was used to draw the bar charts.

3. Results

The occurrence of *Legionella* species, with emphasis on *L. pneumophila,* in the drinking water of the West Bank was determined by cultivation and molecular approaches. Eight sites were sampled over a period of 2.3 years covering two seasonal cycles and the major geographical regions. To gain insight into potential environmental drivers, a broad set of physico-chemical and microbiological parameters were determined. Sampling included bulk water of the DWDS and biofilm analysed by genus-specific plate counts and PCR based methods using environmental DNA extracted directly from the sample material.

3.1 Abundance of *L. pneumophila* in water and biofilm as assessed by cultivation and PCR-based methods

A period of 2.3 years was covered by six sampling campaigns targeting eight drinking water sites associated with local hospitals representative for different regions of the West Bank**.** The six sampling campaigns targeted the main seasons in Palestine, i.e. spring (March to May), summer (June to August), and autumn (October to December). These seasons were sampled twice from 2012 to 2014.

For culture-based assessment of *Legionella* species, 20 biofilm swab samples were analysed per sampling site and sampling campaign, except for the first sampling in autumn 2012 when 44 samples were taken per site to check the variability among samples. DNA was extracted for PCR-based analysis, 5 biofilm swab samples and 1 to 2 representative bulk water samples were taken per sampling campaign that all had been concomitantly analysed by cultivation. Data on sampling per site and sampling campaign are given in detail in supplementary **Table A1**. An overview on the *Legionella* abundances in the bulk water and biofilm is given in **Table 1**.

Table 1: Average *L. pneumophila* abundance per sampling site in water and biofilm as determined by cultivation and PCR.

Sampling site (North to South)	**Coordinates**	**water/ culture: *Legionella* CFU/L ± SD**	**water/ PCR: % *L. Pneumophila* positive ± SD**	**biofilm/ culture: % *Legionella* positive ± SD**	**biofilm/ PCR: % *L. Pneumophila* positive ± SD**
Hospital A (Jenin)	32°27' N, 35°17' E	43.3 (106.1)	66.7 (51.6)	14.6 (20.0)	73.3 (24.2)
Hospital B (Nablus)	32°13' N, 35°14' E	0 (0)	33.3 (40.8)	21.4 (10.8)	63.3 (8.2)
Hospital C (Nablus)	32°13' N, 35°15' E	0 (0)	16.7 (40.8)	2.8 (4.5)	40.0 (25.3)
Hospital D (Ramallah)	31°53' N, 35°12' E	0 (0)	33.3 (28.9)	14.5 (21.0)	66.7 (11.5)
Hospital E (East Jerusalem)	31°46' N, 35°14' E	0 (0)	0 (0)	6.4 (7.2)	60.0 (28.3)
Hospital F (Bethlehem)	31°42' N, 35°11' E	148.0 (229.7)	100.0 (0.0)	29.9 (25.8)	90.0 (11.0)
Hospital G (Hebron)	31°33' N, 35° 04' E	8.3 (20.4)	100.0 (0.0)	23.8 (18.8)	93.3 (16.3)
Hospital H (Hebron)	31°31' N, 35° 05' E	0 (0)	16.7 (40.8)	4.5 (6.4)	60.0 (17.9)
Mean ± *(SD)*		***25.0 (51.9)***	***45.8 (38.6)***	***14.8 (9.8)***	***68.3 (17.3)***

For water samples, *Legionella* plate counts were mostly below detection level, with only three sampling sites out of eight where *Legionella* were detected at all, and only one site with more frequent observation of *Legionella* in summer and autumn (site F). In biofilm, culturable *Legionella* were detected at all sampling sites. Average *Legionella* positive swabs per sampling site ranged from 3 to 30%.

PCR detection of *Legionella* spp. and *L. pneumophila* showed a higher fraction of *Legionella* positive samples. In water samples, at seven out of eight sampling sites *L. pneumophila* was detected with an average detection rate per sampling site ranging from 0 to 100%. In biofilm samples, *Legionella* was regularly detected at all sites at an average detection rate ranging from 40 to 93% per site.

For water samples, *L. pneumophila*-specific PCR was more sensitive than plate counts. The observations by culture and PCR were consistent in a way that whenever plate counts were observed, PCR gave positive results; whereas a large set of PCR-positive samples did not show any plate counts (**Figure 5A**).

Culture based detection of *L. pneumophila* in biofilms was much more successful than in water samples. *Legionella* plate counts from water samples were only positive when about half or more of the biofilm swabs were positive for *L. pneumophila* cultivation (**Figure 5B**).

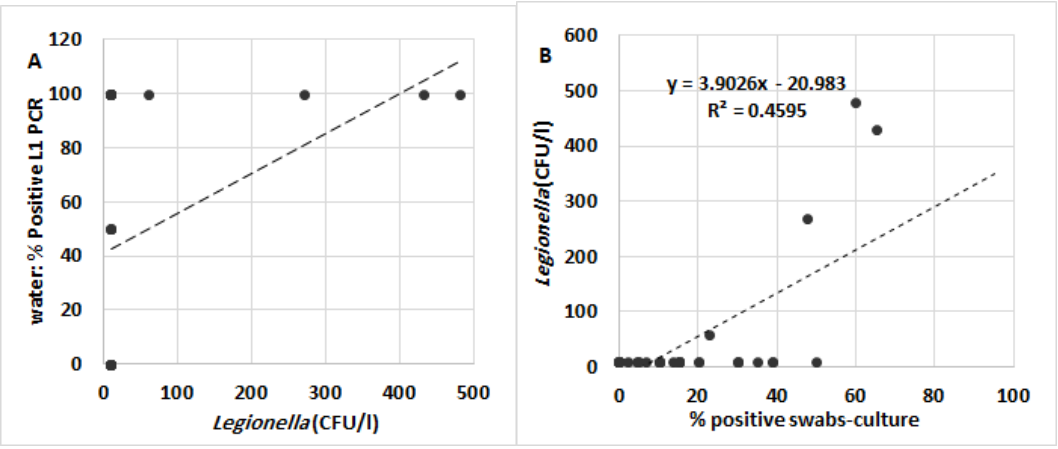


Figure 5: *Legionella* plate counts vs. *L. pneumophila* specific PCR-detection in cold water samples (A), and vs. culture-based positive biofilm swabs (B) (A, B: n = 45)

3.2 Comparison of abundance of culturable *Legionella* in hot and cold water

Hot water was available at five of the eight sampling sites with the exception of spring 2013. At the five sites, sampling was achieved in parallel for hot and cold water. There was no significant difference observed for the *Legionella* plate counts. These counts were usually below detection limit in hot water as in cold water. Only for site F (Bethlehem), there was an increased level of *Legionella* counts in summer 2013 and 2014: in cold water 467 CFU/l in 2013 and 421 CFU/l in 2014; in hot water plate counts were similar (508 CFU/l) in 2013, while in 2014 no *Legionella* were detected. The generally low level of culturable *Legionella* was observed also for the hot water in this sampling comparison (linear correlation: $r^2 = 0.52$, **Figure 6**).

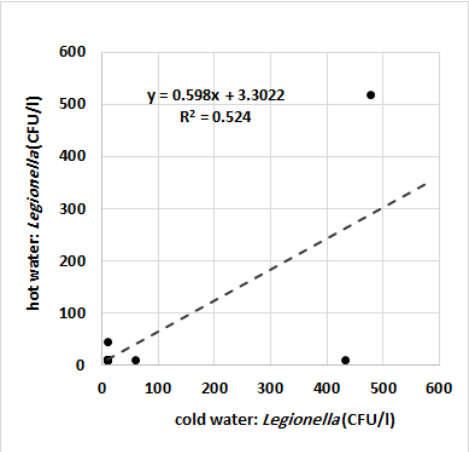


Figure 6: Correlation of *Legionella* plate counts of cold versus hot water (n=24).

3.3 Seasonal dynamics of *L. pneumophila* in biofilm and water

L. pneumophila showed an increase in biofilm samples from spring to autumn across all sampling sites. Both culture based and PCR based methods showed this increase. For PCR based detection, the ratio of positive swabs was in general higher than for culture based detection. Culture based methods and *L. pneumophila* specific PCR based methods showed a good correlation (r^2 = 0.83, correlation with *Legionella* genus specific PCR: r^2 = 0.78, **Figure 7A**). The relative increase was higher for culture along the year (increase from about 10 to 20% of positive swabs) compared to the PCR based detection (increase from about 60 to 80% of positive swabs). The comparison of culture based and PCR based detection of *L. pneumophila* indicated an increase of culturability of *L. pneumophila* from biofilm during the year and the respective exposure to higher temperature from about 15% in spring to about 27% in autumn as shown by the respective ratios (**Figure 7D**).

In contrast, *L. pneumophila* abundance in bulk water showed a maximum in summer as detected by *L. pneumophila* specific PCR detection and by cultivation (**Figure 8A-C**). Detection by PCR was more sensitive than plate counts; plate counts were only recorded when more than 50% of the samples were positive by PCR. The correlation between PCR detection and plate counts (CFU/l) was r^2 = 0.69 (**Figure 8A**). On a seasonal basis, there was no correlation between biofilm and water samples neither by culture nor by PCR based detection. The different seasonal maxima for biofilm and water may be the reason for this discrepancy.

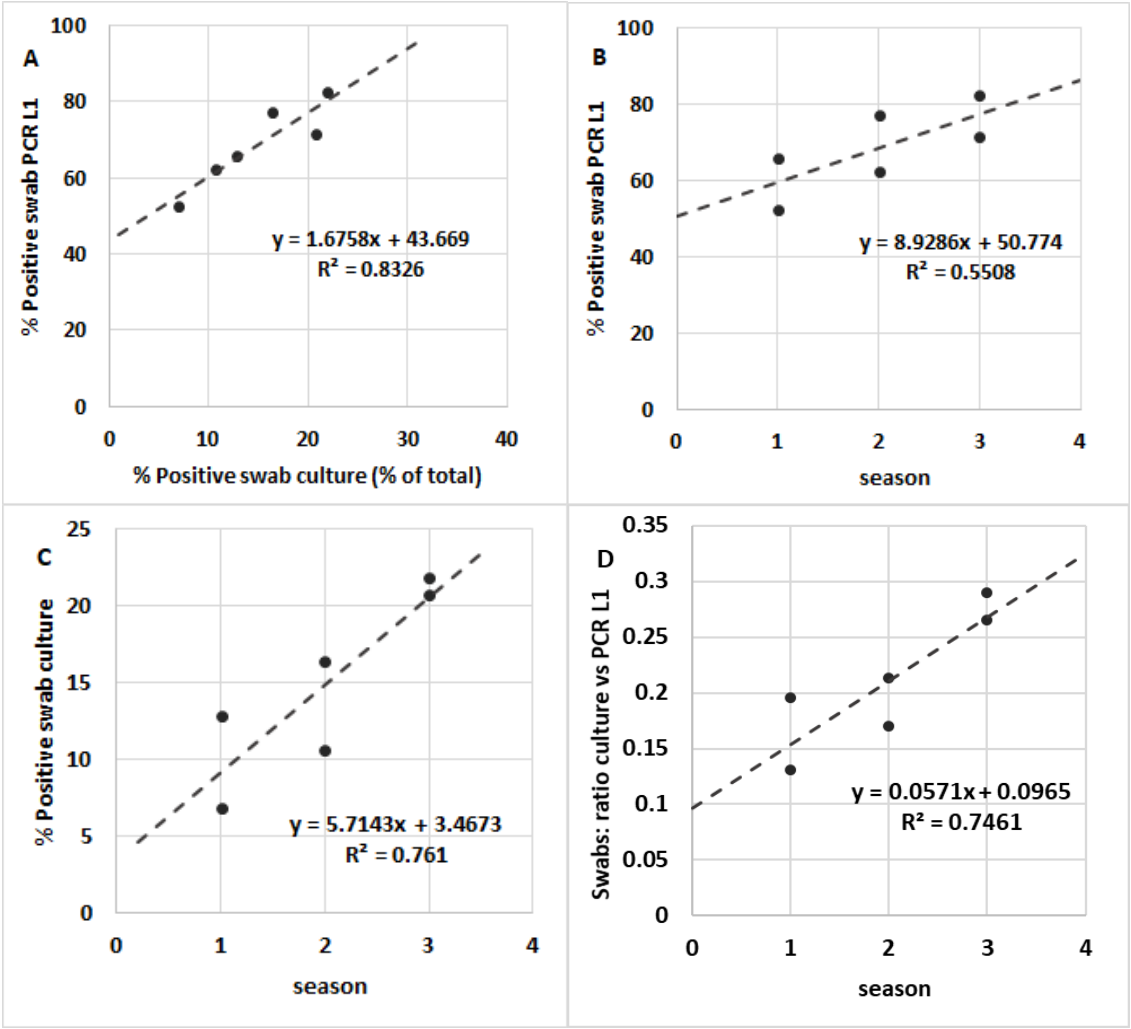


Figure 7: Seasonal variation of abundances of *L. pneumophila* in biofilm of eight sampling sites (average values) of the West Bank sampled from 2012 to 2014. A: swabs positive by culture vs. swabs positive by *L. pneumophila* specific (L1-primer) PCR, B: swabs positive by *L. pneumophila* specific (L1-primer) PCR vs. seasons, C: swabs positive by culture vs. seasons, D: ratio of swabs positive by culture vs. swabs positive by *L. pneumophila* specific PCR vs. seasons (n=45, average n=6). Legend: seasons: 1, spring (March-May); 2, summer (June-September), 3, autumn (October-December).

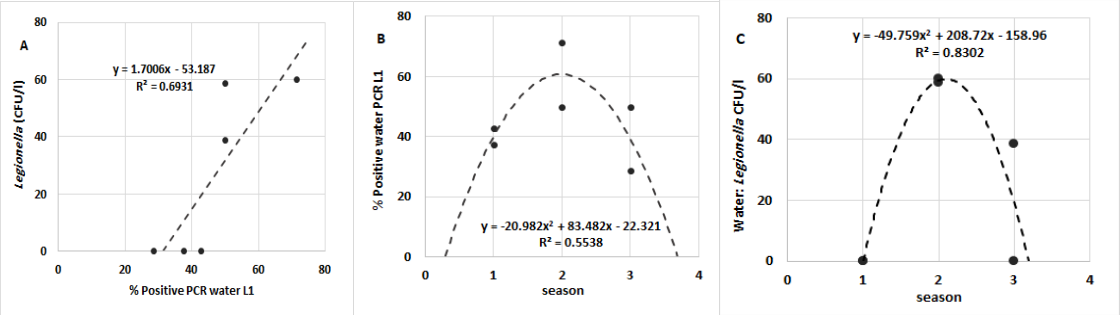


Figure 8: Seasonal variation of abundances of *L. pneumophila* in water of eight sampling sites (average values) of the West Bank sampled from 2012 to 2014. A: *Legionella* plate counts vs. water samples positive by *L. pneumophila* specific (L1-primer) PCR, B: water samples positive *by L. pneumophila* specific (L1-primer) PCR vs. seasons, C: *Legionella* plate counts of water samples vs. seasons (n=45, average n=6). Legend: seasons: 1, spring (March-May); 2, summer (June-September), 3, autumn (October-December).

3.4 Influence of physico-chemical and bacteriological parameters on *Legionella* abundance in water and biofilm

For the assessment of the relationship between *Legionella* abundance in water and biofilm with bacteriological and physico-chemical parameters in water, these parameters were pairwise compared and displayed in a correlation matrix (**Table 2**). 18 quantitatively determined parameters were used to define the physico-chemical background. Heterotrophic plate counts incubated at 25 °C and 37°C were used as general bacteriological parameters. Culturable *Legionella* counts were used for water and biofilm (headline highlighted in yellow). PCR detection of *L. pneumophila* and *Legionella* genus were added for biofilm swabs. For the subspecies level of *L. pneumophila*, the incidence of highly MLVA abundant genotypes and clonal complexes (VACC) were included.

Table 2: Correlation matrix (n=630) of physicochemical and bacteriological parameters in water and biofilm of the eight sampling sites. Data on *Legionella* include culture based and PCR based abundance of *L. pneumophila* and the most abundant genotypes and clonal complexes (VACC).

Sample type	Parameter	HPC-37°C	HPC-25°C	Leg count	Temp	pH	Cond	Iron	Chlorine	Nitrate	Ammonia	Zinc	Turb	HCO3	Chlorides	SO4	Hardness	TDS	Fluorides	Mg	Ca	Ca/Mg Ratio	Gt4(17)	Gt6(18)	Gt10(93)	Gt13(72)	Gt9(92)	Gt10(141)	Gt16(1)	VACC1	VACC2	VACC5	VACC11	Positive swabs-Culture Based CDA (%)	*Leg* spp Primer-CIA (%)	*L.pn* Primer-CIA(%)	* CDA-Culture Based Positive (%)
Water parameter	HPC-37°C	1.000																																		
	HPC-25°C	.942	1.000																																	
	Legionella count	.132	.158	1.000																																
	Temperature	.065	.010	.164	1.000																															
	pH	-.319	-.275	-.287	.082	1.000																														
	Conductivity	-.039	-.073	-.097	-.143	-.312	1.000																													
	Iron	.191	.201	.164	-.142	-.536	.479	1.000																												
	Chlorine	-.137	-.133	-.069	.115	.206	-.403	-.317	1.000																											
	Nitrate	.293	.199	.046	.005	-.216	.115	.357	.045	1.000																										
	Ammonia	.103	.143	-.151	-.123	-.041	.207	.546	-.247	.169	1.000																									
	Zinc	-.159	-.158	-.188	-.123	.092	.017	-.048	.315	.341	.132	1.000																								
	Turbidity	.012	.108	-.035	.137	-.103	.159	.368	-.271	-.030	.287	-.277	1.000																							
	Bicarbonate	.176	.063	.284	.070	-.329	-.071	.158	.084	.146	-.178	-.166	-.514	1.000																						
	Chlorides	-.397	-.460	-.278	-.123	-.025	.425	-.119	.178	-.113	-.177	.009	-.228	.210	1.000																					
	Sulfate	-.033	-.097	-.236	-.085	.057	.081	-.348	.307	.090	-.197	.366	-.790	.251	.440	1.000																				
	Hardness	-.320	-.333	.074	-.085	.225	.027	.102	-.081	-.002	.184	-.097	.038	.010	.153	-.204	1.000																			
	Total Dissolved Solids	-.296	-.287	-.131	-.267	.089	.247	-.159	.178	-.041	.002	.284	-.608	.125	.571	.758	.305	1.000																		
	Fluorides	-.109	-.158	-.079	-.201	-.086	.269	.111	.021	.201	.032	.484	-.104	-.169	.173	.019	-.098	.082	1.000																	
	Magnesium	-.077	-.022	-.508	-.115	.130	.269	.083	-.175	-.064	.287	.091	.526	-.614	.195	-.220	.331	.041	.159	1.000																
	Calcium	-.070	-.169	-.090	-.038	-.122	.156	-.051	.324	.154	-.074	.185	-.539	.438	.617	.643	.113	.666	-.032	-.102	1.000															
	Ca/Mg Ratio	.056	-.022	.499	.110	-.176	-.212	-.048	.262	.135	-.236	-.010	-.598	.668	-.028	.346	-.256	.125	-.150	-.949	.380	1.000														
Water + biofilm parameters	Gt4(17)	.122	.120	.205	-.185	-.192	.112	.161	-.144	-.029	-.058	.070	.165	-.035	.004	-.312	.187	-.124	.302	.215	-.006	-.164	1.000													
	Gt6(18)	-.044	.000	-.038	-.005	.464	-.222	-.317	.630	.038	-.071	.385	-.451	-.003	.096	.509	-.041	.410	.047	-.170	.250	.228	-.138	1.000											
	Gt10(93)	.329	.315	.575	.093	-.183	-.324	.239	-.091	.180	.012	-.155	-.047	.396	-.581	-.315	-.201	-.488	-.127	-.671	-.305	.545	-.065	-.125	1.000										
	Gt13(72)	.129	.139	.176	-.151	-.147	.081	.126	-.127	.006	-.147	.085	.019	.110	-.092	-.204	.171	-.075	.135	.113	.100	-.061	.829	-.106	-.058	1.000									
	Gt9(92)	-.135	-.115	-.085	-.118	.109	-.080	-.077	-.087	-.036	.229	.159	.146	-.302	-.028	-.201	.185	-.034	.346	.292	-.264	-.334	.359	-.099	-.124	.072	1.000								
	Gt10(141)	.063	.086	.852	.228	-.229	-.094	.123	-.042	.035	-.026	-.159	.012	.180	-.307	-.230	-.123	-.264	-.046	-.527	-.261	.474	.009	-.065	.525	-.068	-.064	1.000							
	Gt16(1)	-.014	-.021	.340	-.135	-.200	-.056	.111	-.066	.031	-.115	-.067	-.160	.285	.094	-.035	.467	.337	.027	.005	.375	.119	.513	-.058	-.073	.604	.007	-.038	1.000						
	VACC1	.070	.092	.154	-.182	.073	.011	-.016	.215	.014	-.078	.316	-.108	-.050	.059	.015	.147	.138	.326	.108	.138	-.029	.830	.430	-.140	.682	.297	-.034	.422	1.000					
	VACC2	.097	.157	.315	-.154	-.161	-.011	.106	-.162	-.040	-.254	-.043	-.021	.228	-.238	-.255	.095	-.186	.030	-.132	-.154	.076	.575	-.148	.162	.825	-.009	.143	.431	.426	1.000				
	VACC5	-.029	-.007	.256	-.033	.070	-.028	-.029	-.099	.052	-.247	.053	-.328	.257	.107	.166	.359	.407	.046	-.039	.384	.146	.345	.164	-.122	.507	-.097	-.063	.730	.393	.440	1.000			
	VACC11	.196	.208	.748	.125	-.187	-.271	.176	-.109	.111	.063	-.113	.036	.211	-.517	-.378	-.109	-.438	.021	-.554	-.411	.434	.099	-.145	.839	-.042	.272	.767	-.060	.009	.166	-.142	1.000		
Biofilm parameters	Positive swabs-Culture Based CDA (%)	.043	.087	.678	.123	.040	-.269	-.111	.256	-.028	-.253	.115	-.158	.161	-.200	-.108	.025	-.107	.064	-.411	-.091	.402	.508	.359	.383	.389	.183	.580	.256	.661	.438	.277	.589	1.000		
	Legionella spp Primer-CIA (%)	.247	.315	.184	-.341	-.072	-.237	-.074	.297	-.049	.022	.179	-.248	.071	-.043	.102	-.191	.068	.108	-.247	.114	.282	.234	.305	.237	.130	.129	.127	.090	.372	.161	.042	.265	.464	1.000	
	L. pneumopila Primer-CIA(%)	.321	.355	.275	-.183	-.090	-.216	-.046	.305	-.043	-.070	.136	-.270	.154	-.091	.189	-.206	.110	.022	-.305	.130	.343	.267	.374	.188	.188	.163	.198	.159	.439	.195	.130	.282	.561	.846	1.000	
	* CDA-Culture Based Positive (%)	-.017	.051	.341	-.051	.137	-.368	-.201	.356	-.156	-.270	.139	-.072	.025	-.210	-.144	-.035	-.160	.001	-.306	-.209	.239	.411	.335	.347	.247	.274	.247	.078	.569	.293	.100	.448	.810	.525	.557	1.000

Legend: Color code Red: correlation rs= 0.7-0.9, Orange: correlation rs= 0.5-0.7 and yellow: correlation rs= 0.3-0.5. rs (positive/negative): indicate the positive/negative value of r. *: additional culture based analysis of samples taken for PCR analysis

of *Legionella* in water and biofilm, but also with some genotypes and clonal complexes. Findings of the correlation matrix are elaborated in more detail in the following paragraphs.

As shown in **(Figure 9)**, there was a tight correlation between the Ca/Mg ratio and the Mg concentration for the total data set **(Figure 9A)** as for the average of the eight sampling sites (**Figure 9B**). There was no correlation between Ca and Mg; Ca varied between 103 and 78 mg/l with no correlation with the Mg concentration.

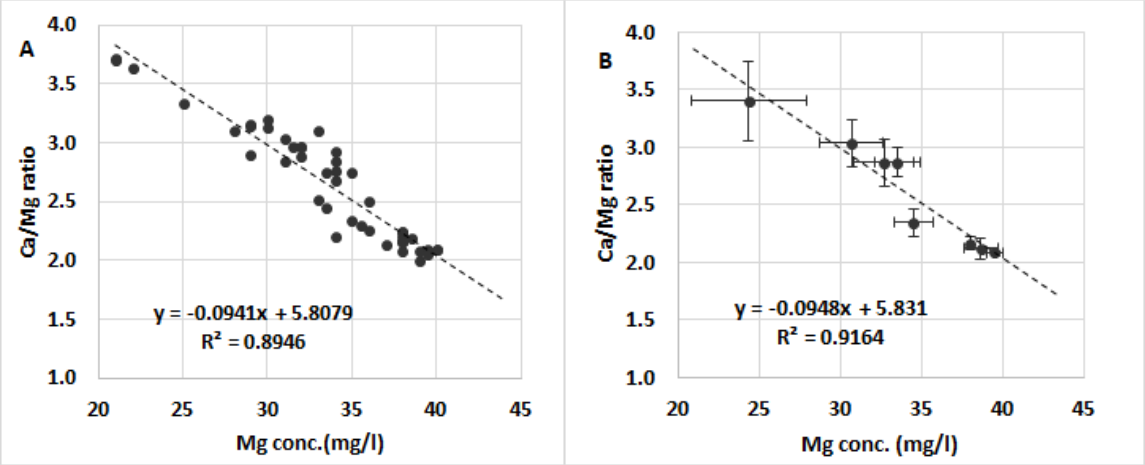


Figure 9: Correlation between the Ca/Mg ratio and the Mg concentration for the total data set (n=45) (Figure 6A) and for the average values of the eight sampling sites (n= 45, average n=8) (Figure 6B).

For the water samples (**Figure 10**), the plate counts of *L. pneumophila* showed a negative correlation with increasing Mg concentrations. This was observed for the overall set of samples as well as for the averages for each sampling site **(Figure 10A)**. For both data sets, the correlation could be best described by a power function with a correlation coefficient of r^2 = 0.54 and 0.78, respectively. **Figure 10B** shows the correlation with the Ca/Mg-ratio that is best described by an exponential function (r^2 = 0.63). Abundance of culturable *L. pneumophila* increased with increasing Ca/Mg ratio as expected from the above described (**Figure 9**) correlation of Mg with the Ca/Mg-ratio. *L. pneumophila*-specific PCR of water samples confirmed the negative correlation of *L. pneumophila* abundance with increasing Mg concentrations (linear correlation of average per sampling site r^2 = 0.49) (**Figure 10C**).

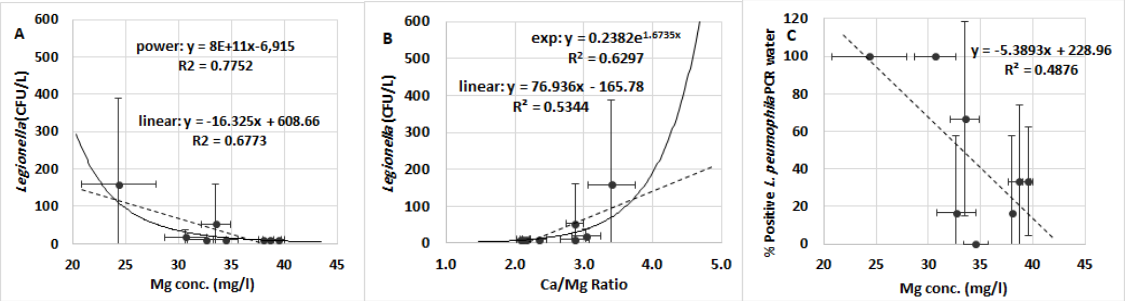


Figure 10: Correlation of *L. pneumophila* abundance with Mg and the Ca/Mg ratio of the water samples as detected by plate counts (A,B) and *L. pneumophila* specific PCR (C) (n=45, average n=8).

For the biofilm samples, a correlation between the magnesium concentration and the PCR based detection of *Legionella* spp. and *L. pneumophila* was observed as shown for the average of each sampling site in **Figure 11**. By contrast, culture based analyses of swabs did not show a clear trend. As for the water samples, increasing Mg concentrations yielded a lower percentage of positive biofilm samples (swabs), both for genus-specific PCR **(Figure 11A)** and *L. pneumophila*- specific PCR **(Figure 11B)**.

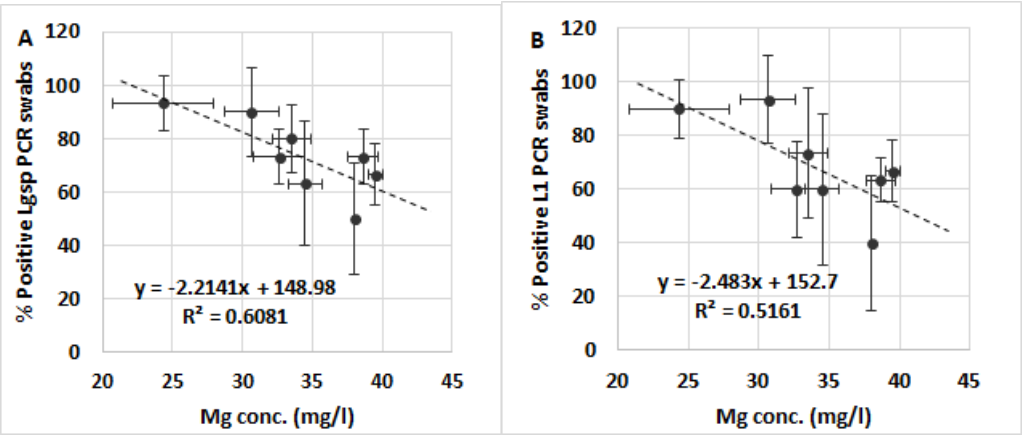


Figure 11: PCR-based detection of the genus *Legionella* (a) and *L. pneumophila* (b) in biofilm swabs vs. magnesium concentration in bulk water (n=45, average n=8).

For culture-based and PCR-based detection of *L. pneumophila* in water and biofilm, correlations with the Ca/Mg ratio yielded similar correlation coefficients as with Mg, but with an increase of Legionella abundance with an increasing Ca/Mg ratio as expected from the correlation between Mg and the Ca/Mg ratio (see **Figure 10** and correlation matrix, **Table 2**).

3.5 Serotype distribution of *L. pneumophila* isolates

Table 3: Serogroup and monoclonal antibody subtyping of 180 environmental *L. pneumophila* isolates from West Bank.

Serogroup	mAb[1] subgroup	*L. pneumophila* isolates	
		No.	Frequency (%)
1	OLDA	10	5.6
1	NA[2]	102	56.7
Total Sg1		*112*	*62.2*
6	Dresden	30	16.7
8		6	3.3
10		1	0.6
(2-14)	NA[2]	31	17.2
Total non Sg1		*68*	*37.8*
Total		180	100

[1]mAb: monoclonal Antibody
[2]NA: Not analyzed

A total of 180 PCR confirmed *L. pneumophila* environmental isolates from eight hospitals across the West Bank and AQU campus were tested for serotyping. The majority of the isolates were characterized as serogroup 1 (Sg.1) (n=112, 62.2%), the remaining isolates (n= 68, 37.8%) were non Sg 1. A subset of 10 Sg1 isolates was subgrouped by monoclonal antibody and all belonged to the MAb 3/1 negative OLDA subtype, which is considered to lack the virulence-associated epitope. The non-Sg.1 isolates (n=37) were analyzed by monoclonal subgrouping and 81% (n=30) of them were serotyped as Sg.6, followed by Sg.8 (n=6) and Sg.10 (n=1). The rest of the non-Sg.1 (n= 31) were characterized as serogroups 2-14 as performed by the agglutination kit **(Table 3).**

In general, the study of the population of *L. pneumophila* in the West Bank showed the following patterns; Sg.1 isolates (62.2%) were more frequent than non-Sg.1 isolates (37.8%). Among the tested isolates, a higher diversity was observed among the non-serogroup1 isolates. In this group the majority of isolates were Sg.6 **(Table 3).**

3.6 Prevalence and biogeography of *L. pneumophila* genotypes across the West bank

From the six sampling campaigns from 2012 to 2014, 197 isolates were obtained. 180 strains were successfully genotyped by MLVA8(12) using 13 loci resulting in 26 different genotypes (ID=0.790, 95% CI 0.739-0.841) (**Table 4 and 5**). 11 genotypes were represented by three strains up to a maximum of 74 strains per genotype (**Table 4**). The remaining 15 genotypes were represented by 2 or 1 strain. The 26 genotypes are affiliated with the 4 VNTR clonal complexes VACC 1, 2, 5, and 11, indicating the relatedness among the genotypes. VACC1, VACC2 and VACC5 were clonal complexes previously defined at the MLVA *Legionella* database. However, VACC11 is described for the first time in this study **(Figure 12 Figure A1).** 17 of the MLVA genotypes are considered to be affiliated with 12 different sequence types (ST), meaning that some of the MLVA genotypes pertained to the same ST. In the following analyses, the focus will be on the genotyped strains; genotyping will be used as a classification scheme for further studies of the ecology of *L. pneumophila.*

Table 4. Sampling location and number of isolates obtained from each site from six sampling campaigns in 2012-2014.

Sampling site (North to South)	Coordinates	No. of isolates	Percent of total isolates (%)	No. of genotypes	Ratio: genotypes/ isolates
A Jenin	32°27' N, 35°17' E	28	15.6	5	0.19
B Nablus	32°13' N, 35°14' E	32	17.8	5	0.16
C Nablus	32°13' N, 35°15' E	5	2.8	2	0.4
D Ramallah	31°53' N, 35°12' E	18	10	2	0.11
Al-Quds University/ East Jerusalem	31°45' N, 35°15' E	15	8.3	6	0.4
E East Jerusalem	31°46' N, 35°14' E	10	5.6	6	0.6
F Bethlehem	31°42' N, 35°11' E	32	17.8	7	0.22
G Hebron	31°33' N, 35° 04' E	34	18.9	5	0.15
H Hebron	31°31' N, 35° 05' E	6	3.3	4	0.67
Total		***180***	***100***	***26***	

Though sampling was achieved in the same way for seven hospital drinking water sampling sites, the yield of isolates per site was rather variable **(Table 4)**. It should be noted that sampling site D spring, summer and autumn could only be sampled once, while all other seven sites were sampled twice for these seasons (see Supplementary Material, **Table A1**).

When analyzing the MLVA-8(12) genotypes, the same main clusters were observed and only minor changes were detected. By using the combination of 13 loci, a small group of six isolates that belonged to two different genotypes separated from the large VACC1 and remained as singletons. These six isolates were all Sg.8 strains isolated from AQU in contrast with the rest of the strains enclosed in VACC1, which were all Sg.1. They differed from the rest of the isolates contained in VACC1 in the number of repeats observed for VNTR markers Lpms31, which presented 17 repeats in comparison to 4 or 0 in the rest of the profiles of VACC1, and VNTR Lpms33 and Lpms34, which presented both only one repeat in contrast to 4 and 2 repeats found, respectively, in VACC1.In total, 96.6% of isolates (n=174) were clustered in the four VACCs and the rest 3.6% (n=6) were found as singletons. VACC1 was the largest cluster including 110 isolates (61.2%). VACC11, VACC2 and VACC5 were in comparison small clusters counting 31 (17.2%), 19 (10.5%) and 14 (7.7%) isolates respectively. Only one Sg.1 isolate (A100) isolated from biofilm in hospital F was as well included in VACC2. According to the MLVA genotyping, the West Bank is characterized by the presence of few singletons, most of their genotypes were isolated more than once. Additionally, the diversity was (ID=0.790, 95% CI 0.739- 0.841) in the West Bank although the sampled area extended along the West Bank **(Figure 12).**

Table 5: MLVA-Genotype composition of isolated *L. pneumophila* strains and their affiliation with VNTR clonal complexes (VACC) and sequence types (ST).

MLVA-8(12) genotype	VACC	No.of strains	Frequency (%)	Sequence Type (ST)
Gt4(17)^{AQ}	1	74	41.1	ST 1
Gt6(18)$^{w(G)}$	1	30	16.7	ST 1*
Gt10(93)$^{w(F)}$	11	16	8.9	ST 461
Gt13(72)	2	10	5.6	ST 1326
Gt9(92)	11	8	4.4	ST 461
Gt10(141)$^{W(F)}$	11	6	3.3	ST 461*
Gt12(84)^{AQ}	1	5	2.8	ST 1358
Gt16(1)$^{w(A)}$	5	5	2.8	ST 1438
Gt40(47)	5	3	1.7	ST 292*
Gt63(83)	1	3	1.7	NA
Gt64(74)	2	3	1.7	ST 9*
Gt13(143)	2	2	1.1	ST 1326^{e}
Gt8(7)	5	2	1.1	ST 1482
Gt11(87)^{AQ}	1	1	0.6	ST 1358
Gt13(106)^{AQ}	2	1	0.6	ST 1326^{e}
Gt16(3)	5	1	0.6	ST1438^{e}
Gt16(6)	5	1	0.6	ST1438^{e}
Gt24(68)	2	1	0.6	ST 93*
Gt38(109)	2	1	0.6	NA
Gt4(16)	1	1	0.6	ST 1*
Gt4(20)^{AQ}	1	1	0.6	ST 1*
Gt55(94)	11	1	0.6	NA
Gt6(15)	1	1	0.6	ST 1*
Gt8(142)	5	1	0.6	ST1482^{e}
Gt8(23)	5	1	0.6	ST1482^{e}
Gt84(106)^{AQ}	2	1	0.6	ST 187*
Total No. 26	**4**	**180**	**100**	**12 ST's vs. 23 Gts**

Legend: NA, not available; *, ST was assessed for strains of the same MLVA-8(12) genotype, and not directly for the West Bank strains; ^{e}, ST was estimated from the MLVA-8 pattern; ^{W}, genotype retrieved from water, in brackets the site of isolation is indicated; ^{AQ}, contains strains retrieved from biofilm of the AQU.

The majority of isolates were retrieved from biofilm swabs (192 strains in total with 175 genotyped strains) and only a minor fraction from water samples (5 strains in total, 3 from Bethlehem (sampling site F), one from Jenin (sampling site A) and Hebron (sampling site G)). The genotypes isolated from water are indicated in **table 5**; these genotypes are among the most abundant genotypes in general, and of high relevance for the respective sampling site. For a comparison, the 15 isolates retrieved from biofilm of the Al Quds-University are added, and indicated in the strain table and the biogeographic distribution (**Table 4, 5 and Figure 13**).

The biogeographic distribution of the strains according to their MLVA-genotype and clonal VNTR complex (VACC) is indicated in **Figure 13**. On the genotype level, **Figure 13A** shows a genotype pattern that varies on the regional level. In the north of the West Bank, genotype 4(17) is highly abundant. In the south the pattern shows high divergence from site to site. For example, site G is dominated by genotype 6(18) that was not retrieved from any other site; similarly site F showed a high prevalence of Gt 10(141) that was endemic for this site. Also on the level of the clonal complexes the regional variability was well pronounced (**Figure 13B**). In general, there is a high prevalence of VACC1, except for sites E, F and H in the south. Site F showed a high prevalence of VACC11, site E for VACC 2 and site H for VACC 5.

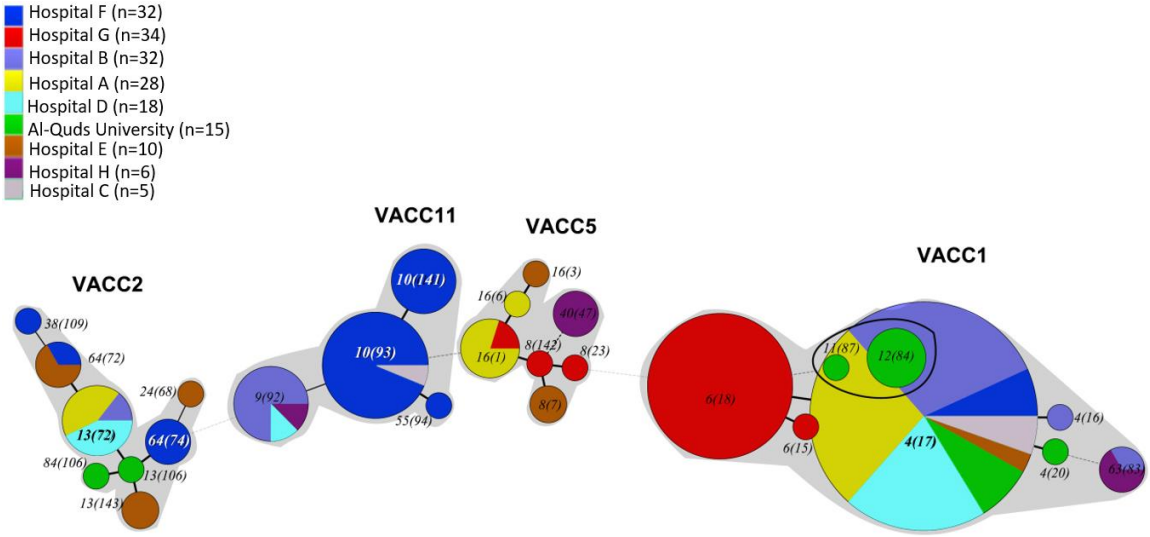


Figure 12: Minimum-spanning tree based on MLVA-8(12) profiles of 180 *L. pneumophila* strains isolated from the West Bank. Each circle in the tree represents a different MLVA-8(12) genotype. The genotype number is indicated within or near the circle, which is proportional to the genotype frequencies. Different colors in the pie charts refer to the eight sampling locations (see legend). Thickness of the branches represents the number of different loci. MLVA clonal complexes (VACC) have been circled. The circles representing the Sg6 singletons from Al- Quds University (Gt11 (87) and Gt12(84)) overlap visually within the circle that represent Gt4(17) due to the high abundance of this genotype

In particular, the most frequently isolated genotype was Gt4 (17) (n=74, 41.1%), followed by Gt6 (18) (n=30, 16.7%). Both of them corresponded to Sequence Type 1 (ST1) (39). The following most frequent genotypes were Gt10 (93) (n=16, 8.9%), and Gt9 (92) (n=8, 4.4%), corresponding to ST461, and Gt13 (72) (n=10, 5.6%), which corresponded to ST1326 **(Table 5 and Figure A1)**.

At the genotype level, only eight out of the 26 MLVA-8(12) genotypes were isolated in more than one location. The rest 18 MLVA-8(12) genotypes were isolated exclusively in a certain site. Gt4 (17), known to correspond to the highly broad distributed ST1, was the only genotype present in all hospitals except in hospital G. Furthermore, it was not only present but represented a high fraction of the isolates in several hospitals. Gt4 (17) was the most abundant genotype in hospital B (n=22, 68.7%), A (n=20, 71.4%), D (n=15, 83.4%), and C hospitals (n=4, 80%). At Al-Quds University it accounted for 40% of the genotypes (n=6). Gt6(18), which only differed in one repeat in the VNTR Lpms35 and also corresponded to ST1, was found exclusively in hospital G. It was the most abundant genotype at hospital G (n=30, 90.9%). Genotype Gt10 (93), which corresponded to ST461 was found in hospitals C and F, yet, it was isolated only once (n=1, 20%) in hospital C and in contrast was the most abundant genotype in hospital F (n=15, 45.5%).Other genotypes that were found in more than one location, usually only in one or two locations, can be observed in **(Figure 12 and Table 6).**

Table 6 (Part 1): MLVA-8(12) genotypes abundance in the West Bank

Location	MLVA-8 (12)	No. Genotype (%)	Sg-MAb	MLVA_CC	No. of VACCs (%)
Hospital A	Gt4(17)	20 (71)	1	VACC1	20(71)
	Gt16(1)	4 (14)	6 Dresden & (2-14)	VACC5	5(18)
	Gt13(72)	3 (11)	(2-14)	VACC2	3(11)
	Gt16(6)	1 (4)	(2-14)	0	0
	Total	*28(100)*	*0*	*0*	*28(100)*
Hospital B	Gt4(17)	21(66)	1	VACC1	24(75)
	Gt9(92)	7(22)	(2-14)	VACC11	7(22)
	Gt63(83)	2(6)	1	VACC2	1(3)
	Gt4(16)	1(3)	1	0	0
	Gt13(72)	1(3)	(2-14)	0	0
	Total	*32(100)*	*0*	*0*	*32(100)*
Hospital C	Gt4(17)	4(80)	1	VACC1	4(80)
	Gt10(93)	1(20)	(2-14)	VACC11	1(20)
	Total	*5(100)*	*0*	*0*	*5(100)*
Hospital D	Gt4(17)	15(83)	1	VACC1	15(83)
	Gt13(72)	3(17)	(2-14)	VACC2	3(17)
	Total	*18(100)*	*0*	*0*	*18(100)*
Hospital E	Gt13(143)	2(20)	(2-14) & Sg.10	VACC2	5(50)
	Gt64(72)	2(20)	6 Dresden & (2-14)	VACC5	3(30)
	Gt8(7)	2(20)	(2-14)	VACC1	2(20)
	Gt4(17)	2(20)	1	0	0
	Gt24(68)	1(10)	(2-14)	0	0
	Gt16(3)	1(10)	(2-14)	0	0
	Total	*10(100)*	*0*	*0*	*10(100)*

Table 6 (part 2): MLVA-8(12) genotypes abundance in the West Bank					
Hospital F	Gt10(93)	14(44)	6 Dresden & (2-14)	VACC11	21(66)
	Gt10(141)	6(19)	6 Dresden	VACC1	6(19)
	Gt4(17)	6(19)	1	VACC2	5(16)
	Gt64(74)	3(9)	6 Dresden	0	0
	Gt55(94)	1(3)	6 Dresden	0	0
	Gt64(72)	1(3)	6 Dresden	0	0
	Gt38(109)	1(3)	1	0	0
	Total	*32(100)*	*0*	*0*	*32(100)*
Hospital G	Gt6(18)	30(88)	1	VACC1	31(91)
	Gt6(15)	1(3)	1	VACC5	3(9)
	Gt16(1)	1(3)	(2-14)	0	0
	Gt8(142)	1(3)	(2-14)	0	0
	Gt8(23)	1(3)	(2-14)	0	0
	Total	*34(100)*	*0*	*0*	*34(100)*
Hospital H	Gt40(47)	3(50)	6 Dresden	VACC5	3(50)
	Gt10(93)	1(17)	6 Dresden	VACC11	2(33)
	Gt9(92)	1(17)	(2-14)	VACC1	1(17)
	Gt63(83)	1(17)	1	0	0
	Total	*6(100)*	*0*	*0*	*6(100)*
AQU*	Gt4(17)	6(40)	1 OLDA	VACC1	13(87)
	Gt12(84)	5(33)	8	VACC2	2(13)
	Gt4(20)	1(7)	1 OLDA	0	0
	Gt11(87)	1(7)	8	0	0
	Gt13(106)	1(7)	6	0	0
	Gt84(106)	1(7)	6	0	0
	Total	*15(100)*	*0*	*0*	*15(100)*
AQU*: Al-Quds University					

In **Figure 14**, the Venn diagram shows the number of MLVA8(12) genotypes shared among the North, Central and Southern West Bank. According to the West Bank geographical distribution, 64(35.6%), 43(23.9%) and 73(40.6%) isolates were isolated from Northern, Central and Southern West Bank respectively. Nevertheless, the most abundant and more broadly distributed genotype were common to the whole West Bank, in particular, Gt4(17), which constituted 68.8%, 53.5% and 9.6% of total isolates in Northern, Central and Southern WB, respectively. Gt13 (17) was shared between Northern and Central West Bank. Gt64 (72) shared between Central and Southern WB. Surprisingly, four genotypes [Gt9 (92), Gt10 (93), Gt16 (1) and Gt63 (83)] were shared between Northern and Southern WB which are geographically separated. The least diversity of the genotypes were in the Northern WB. Whereas, the diversity was slightly higher in Southern than Central West Bank.

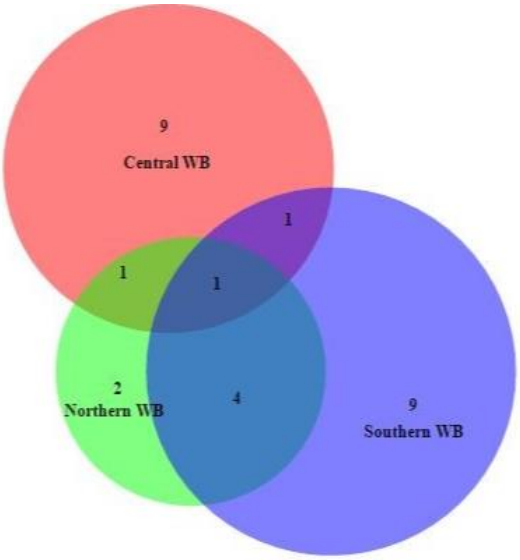


Figure 14: Venn diagram of the number of MLVA-8(12) genotypes shared among distinct geographic areas. The size of the circle is proportional to the number of genotypes of each area of the West Bank, Palestine.

The richness, i.e. the number of *L. pneumophila* MLVA genotypes, varied from 2 to 7 per sampling site, with an average value of 4.5 genotypes for the eight sampling sites. The ratio of number of genotypes vs. number of strains retrieved per sampling site was added as an indicator of the "genotype diversity" **(Table 4)**, ranging from 0.11 to 0.67 and on average of 0.30. There was no significant correlation (r^2 = 0.28) between the number of strains and the number of genotypes retrieved per sampling site. Correlation analysis revealed a correlation between the average percentage of biofilm swabs positive for *Legionella* culture per sampling site and the "genotype diversity" (**Figure 13C**) sites with low *Legionella* incidence on biofilm swabs showed a high diversity compared to a low diversity in case of high *Legionella* incidence.

Correlation analyses with the number of genotypes, i.e. the richness, per site vs. the abundance of culturable *Legionella* in biofilm swabs or water or vs. the PCR detection of *L. pneumophila* on biofilm swabs showed rather weak correlations (r^2 < 0.4). No obvious correlations occurred between the genotype diversity and the PCR detection of *L. pneumophila* in biofilm swabs (r^2 = 0.27) or the detection of culturable *Legionella* in water (r^2 = 0.06).

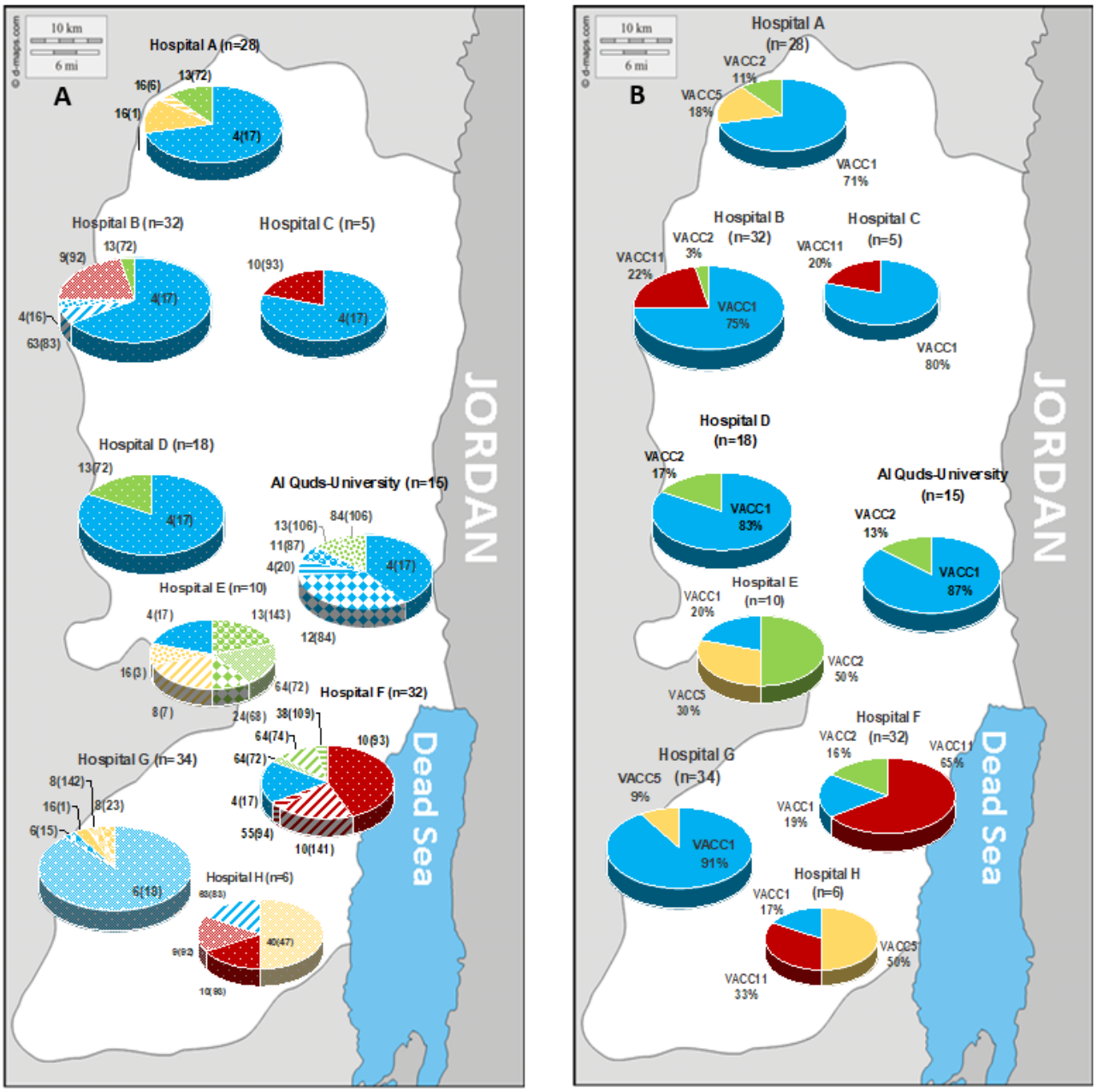


Figure 13: Biogeographic distribution of the *L. pneumophila* strains according to their MLVA-8(12) genotype (A) and their clonal complex (VACC) (B). In **Figure 13A** the respective MLVA-8(12) genotype is indicated in the following way: genotype MLVA-8 plus MLVA-12 in brackets, e.g. the genotype MLVA-8(12) 4(17) is indicated as "4(17)".

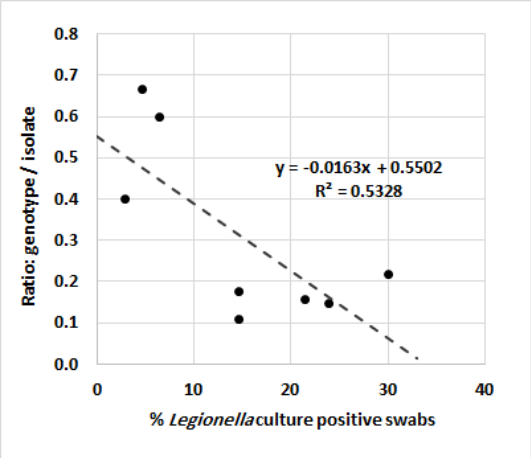


Figure 13C: shows the genotype diversity (number of genotypes / number of strains retrieved per sampling site) vs. the average percentage of *Legionella* positive biofilm swabs (by cultivation) per sampling site (n=45, average n=8).

Table 7. VNTR characteristics of the *L. pneumophila* strains isolated from the West Bank

VNTR	West Bank		
	No. of repeats	HGDI[1] (CI 95%)	Null alleles (%)
Lpms1	4	0.528 (0.459-0.596)	0
Lpms3	2	0.461 (0.420-0.502)	0
Lpms13	5	0.579 (0.506-0.652)	0
Lpms17	2	0.115 (0.053-0.178)	0
Lpms19	2	0.022 (1.000-0.053)	1.11
Lpms31	6	0.576 (0.513-0.639)	1.67
Lpms33	4	0.575 (0.506-0.643)	0
Lpms34	4	0.503 (0.429-0.577)	2.22
Lpms35	6	0.687 (0.641-0.733)	1.67
Lpms38	3	0.249 (0.168-0.330)	4.44
Lpms39	3	0.509 (0.445-0.574)	0
Lpms40	3	0.493 (0.444-0.541)	3.33
Lpms44	3	0.498 (0.463-0.533)	0

[1] HGDI: Hunter-Gaston Discrimination Index

A more detailed analysis of the population structure at the level of the VNTR markers that configure the genotypes showed that, overall, balance variability in the number of repeats for most VNTR markers was observed among the isolates **(Table 7).** This could be due to the homogenized habitat and location where the isolates have been obtained from. Some VNTR markers appeared to be less variable and showed a reduced number of repeats, as shown for markers Lpms3, Lpms17 and Lpms19. However, other markers showed a greater variability as Lpms31 and Lpms35. In general, the same repeats prevailed independently of the area where the isolates were isolated. Null alleles were present at different frequencies in distinct VNTR markers. Especially high were the frequencies of null alleles in Lpms38. Remarkably, a new allele of Lpms34 was described during this study. The new allele had a size of 634 base pairs and was formed of four repeats. A total of 31 (17.2%) isolates contained this allele. However, it was very infrequently found in the West Bank. This allele was not described before in previous MLVA studies for *L. pneumophila* and therefore should be incorporated in the allele assignment reference table provided at *L. pneumophila* MLVA database for future studies **(Table 7)**.

3.7 Environmental factors correlating with genotype abundance and composition

The variable pattern of genotype prevalence and composition raised the question concerning the influencing factors. There was a broad set of physico-chemical parameters recorded or obtained from the water authorities of the West Bank (**Table A1**, Supplementary Material). The quantitavely measured parameters that showed differences between the sampling sites were used for detailed statistical analysis, i. e. PCA and cluster analysis. The selection criterion for the genotypes to be included in these analyses was the number of strains available per genotype, i.e. only genotypes represented by at least three isolates were included. This selection resulted in ten genotypes submitted to PCA and cluster analyses.

Three groups of genotypes could be separated by PCA and cluster analysis based on the environmental parameters, i.e. concentrations of chloride, sulphate, Mg, Ca/Mg-ratio ("Ratio"), Ca, total dissolved solids (TDS), and turbidity and *Legionella* plate counts **(Figure 15).** Cluster analyses formed three groups for the ten genotypes. By PCA, eight of the ten genotypes were assigned to three groups, with two genotypes (Gt 9(92) and Gt 63(83)) being close the PCA-group B1 comprising Gt 4(17) and Gt 13(72). These four genotypes (Gt 9(92), 63(83), 4(17), 13(72) were included in one group (cluster group B) by cluster analyses. For further analyses and considerations, these four genotypes are treated as members of group B.

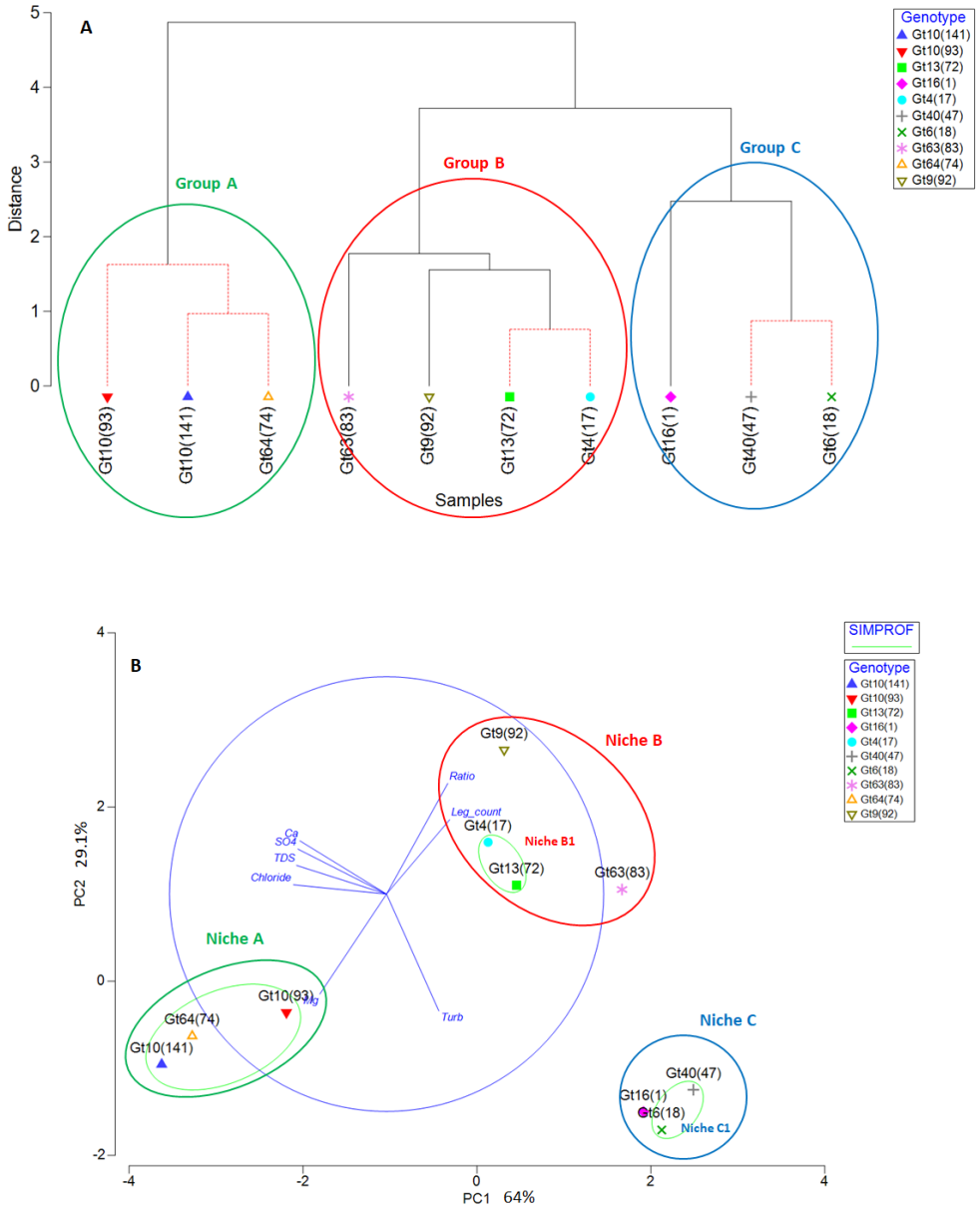


Figure 15: A) Dendogram showing group average hierarchical clustering of MLVA-8(12) genotypes. **B)** Principal-component analysis (PCA) comparison of MLVA-8(12) genotypes showing the effect of biological and physicochemical parameters. Light green ellipses represent the log normal distributions of principal component values for genotype groups. Close genotype groups were summarized as pertaining to the same niche. The resulting niches are A (green), niche B (red), niche C (blue). Niche A and C are consistent with the calculated grouping A and C; group B1 was enlarged to include two more genotypes in group B that concomitantly represents niche B. Legend: Ratio, Ca-Mg ratio; Leg. count, *Legionella* plate counts in water samples.

Table 8 provides the specific characteristics of the niches A, B, and C in terms of summary of the parameters, their concentrations and distinction per parameter. It demonstrates that the group of strains have clearly distinct niches. For more details on statistical analyses see **Table A2**.

Table 8: Summary of environmental parameters describing the different niches of *L. pneumophila* MLVA-genotypes and their significant differences based on a correlation analysis

Niche design-ation	VACC: genotypes (Gt)	No. isol-ates	Statist ics	*Legion-ella* (CFU/l)	Turb (mg/ l)	Chlori-de (mg/l)	SO4 (mg/l)	TDS (mg/l)	Mg (mg/l)	Ca (mg/l)	Ca/Mg ratio
A	VACC11: Gt10(93), Gt10(141) VACC2: Gt64(74)	25	Mean	263	1.3	27.6	13.8	284.3	23.8	80.5	3.4
			SD	235	0.2	6.7	3.8	5.9	4.2	2.9	0.4
B	VACC1: Gt4(17), Gt63(83) VACC2: Gt13(72), VACC11: Gt9(92)	95	Mean	61	1.3	52	14.9	314.3	34.2	81.8	2.3
			SD	122	0.4	21.1	5.6	91.8	9.2	21.2	0.7
C	VACC1: Gt6(18) VACC5: Gt16(1), Gt40(47)	38	Mean	41	0.9	62.3	32.7	405.7	31.2	93.8	3
			SD	79	0.1	4.7	6.2	37.8	2.1	3.8	0.2
A vs. B			*p**	***	NS	***	**	**	***	*	***
A vs. C			*p**	***	***	***	***	***	***	***	**
B vs.C			*p**	NS	***	**	***	***	**	***	***

Legend: Independent t-test: NS, not significant; *, *p≤ 0.05; **, p≤ 0.01; ***, p≤ 0.001.*

The environment of these three groups is characterized by the respective parameters in summary and assigns distinct niches to the three groups. The characteristisation of the three groups by the parameters is shown in **Table 8**. While a comprehensive and stable distinction is shown for the summary of the parameters, also many of the single parameters allow a distinction between two or three groups.

These three groups of genotypes were considered to co-occur in their respective environment as described by the above mentioned parameters. The respective habitats as described by these environmental parameters for the three groups of genotypes could be considered as their “niche”. This means that to each niche three to four genotypes were assigned. More genotypes maybe sharing these niches, but they were not included in the analysis due to a low strain number. For example, genotypes Gt9(92) and Gt63(83) were combined with Gt4(17) and Gt13(72) in niche B, because they can be considered to live in a comparable environment.

3.8 Relationship between genotype presence and abundance of *L. pneumophila* in water and biofilm

Since the overall *Legionella* abundance is critical for environmental issues and human health, its relationship with genotype abundance was analysed in more detail. Such a relationship was also indicated by the PCA and cluster analyses of the MLVA genotypes **(Figure 6)**. Per sampling site, the percentage of culturable *L. pneumophila* positive biofilm samples correlated well with the presence of strains of the clonal complex VACC1 (r^2 = 0.53), and was improved when adding the locally highly relevant strains of VACC11 (r^2 = 0.86) (**Figure 16A,B**). VACC11 alone did not show a good correlation with the biofilm (r^2 = 0.18), due to their endemic distribution.

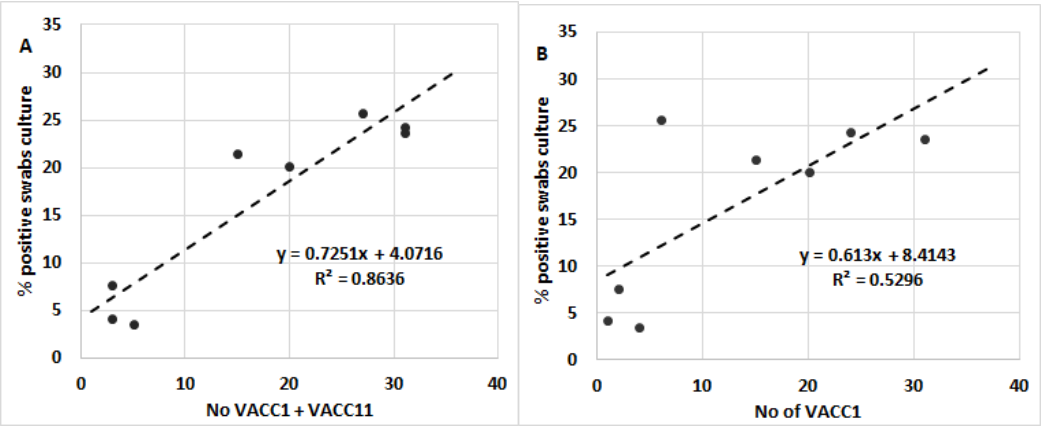


Figure 16: Biofilm samples: Correlation of *Legionella* culture positive biofilm samples was obtained vs. the abundance of strains of the clonal complexes VACC1 + VACC11 (n=135, average n=8) (A); for a comparison, the correlation vs. VACC1 alone is shown (n=103, average n=8) (B).

The *Legionella* plate counts obtained from water samples, correlated best in terms of clonal complex with VACC11 **(Figure 17)**. In terms of correlation with genotypes, best correlation was obtained with the abundance of the VACC11 genotype Gt 10(141), followed by 10(93). As combination, the VACC11-genotypes 10(141) and 10(93) showed the best correlation (**Figure 17D**). However, this observation was strongly influenced by the high *Legionella* counts from site F and the respective genotypes present. Indeed, the three isolates retrieved from water samples of site F were identified as two strains of the VACC11-genotype 10(93) and one strain of VACC11-genotype 10(141). Please note that for the calculations based on genotype abundance the whole set of isolates consisting of 175 biofilm-derived strains and 5 water-derived strains was used.

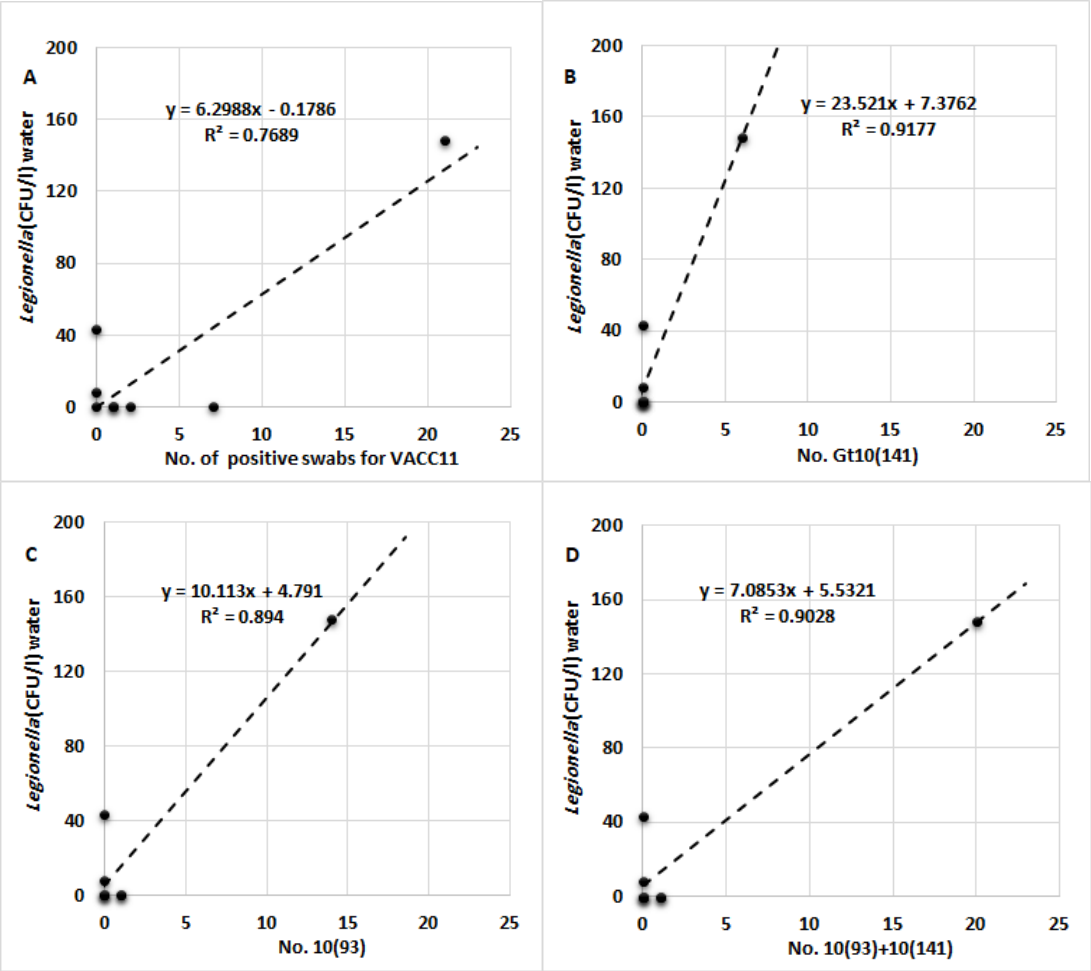


Figure 17: Water samples: Clonal complex (n=32) (A) and individual genotypes (n=6 and n=16, respectively) (B, C) and combination of genotypes (n=22) (D) correlating best with *Legionella* plate counts obtained from water samples.

3.9 Influence of magnesium concentration on abundance of genotypes and clonal complexes

Figure 18 shows the details on the relationship between the magnesium concentration and the relevant clonal complexes and genotypes. For an overview, please refer to the correlation matrix (**Table 2**). There was only a pronounced correlation between VACC11 and the magnesium concentration. All other clonal complexes including the highly abundant VACC1 did not show a significant correlation (**Figure 18D**). VACC11 showed the highest abundance at low Mg concentrations. The occurrence at lower Mg concentrations were even more pronounced for Gt10(93) and most pronounced for Gt10(141). There was no obvious pronounced relationship with Mg for the highly abundant genotypes. For low abundant genotypes the data did not allow an analysis.

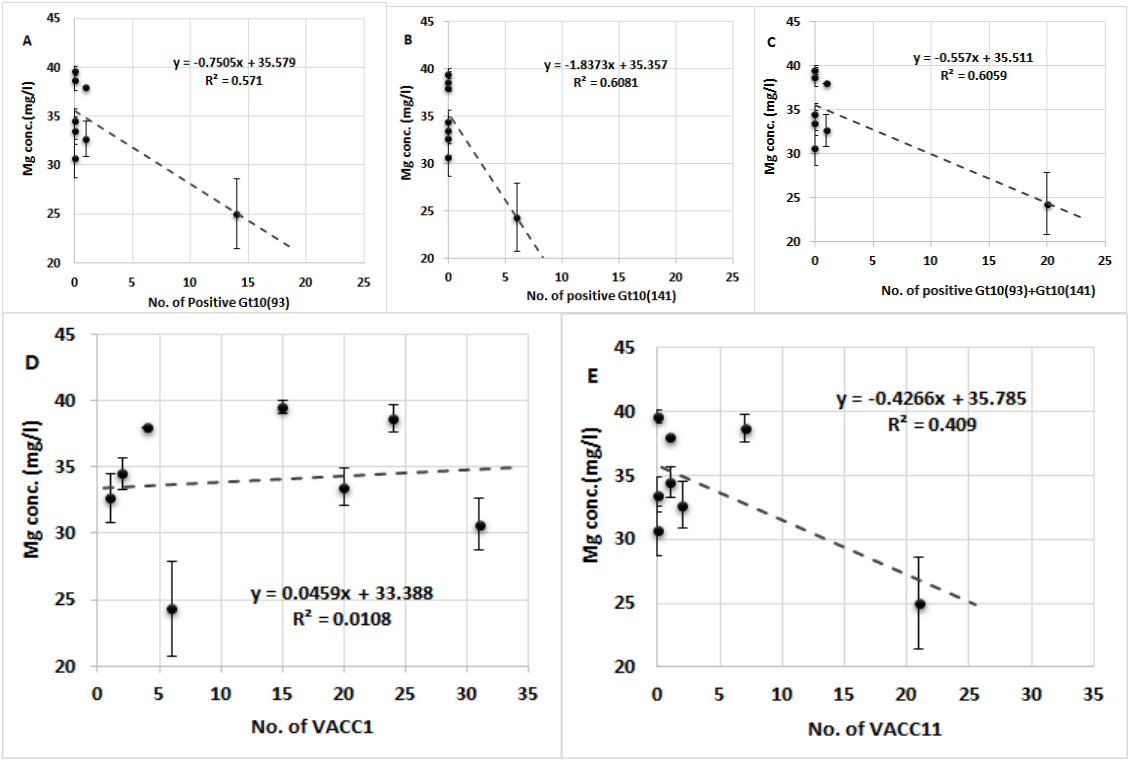


Figure 18: Correlation between the Mg concentration and the abundance of the relevant genotypes Gt10(93) and Gt 10(141) (n=16, n=6 and n=22, respectively) (A-C, individual and combined) and clonal complexes VACC1 (n=103) (D) and VACC11 (n=32) (E) at the respective sampling sites.

3.10 Comparison of virulence activity of *L. pneumophila* reference strains

Three *L. pneumophila* reference strains (Paris, Corby and Philadelphia-1) were tested for their virulence effect on *A. castellanii*, THP-1 macrophages and pore forming mediated cytotoxicity in which AA100 wild type *L. pneumophila* and dotA mutant were used as control strains.

L. pneumophila strain Paris which is Gt4 (17) included in VACC1 a large cluster in the West Bank and worldwide showed the highest virulent activity for cytopathogenicity of THP-1 macrophages, pore-forming mediated cytotoxicity and Infection with *A. castellanii*. *L. pneumophila* strain Corby and WTAA100 have nearly the same high virulent activity, since WTAA100 is *L. pneumophila* strain Corby. Also, *L. pneumophila* strain Philadelphia-1 included in VACC2 have the least virulent activity in comparison with the previous two *L. pneumophila* strains. However, the non-virulent dotA mutant resulted in less than 15% cell death **(Figure 19).**

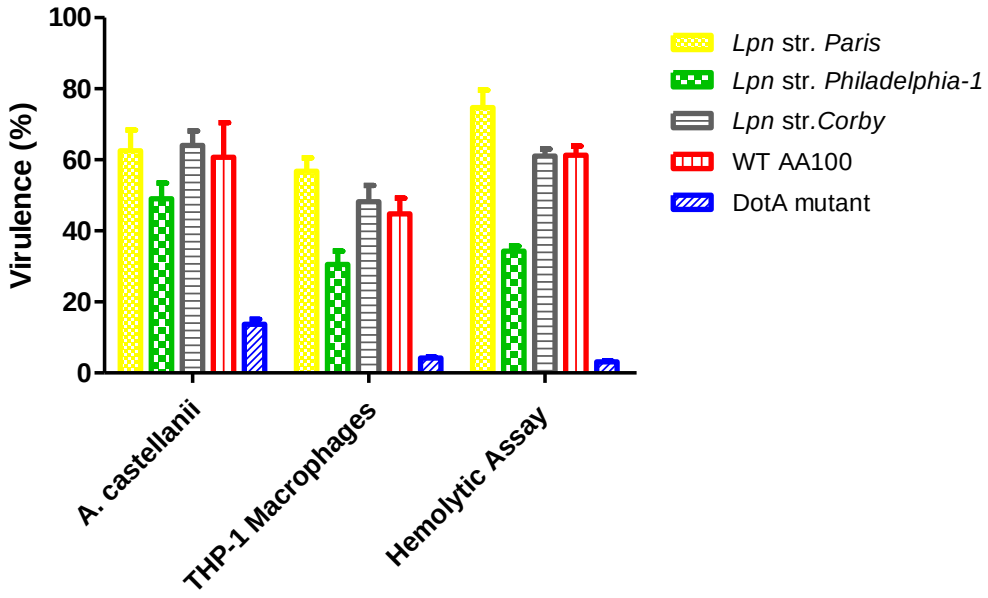


Figure 19: The virulent activity of *L. pneumophila* reference strains during post exponential phase was determined by infectivity of *A. castellanii* and THP-1 macrophages and pore forming mediated cytotoxicity of sRBC's . Each test was done in triplicate.

One sample t-test was performed to show the difference between the reference strains for each test independently. Statistically significant differences were shown for Infection with *A. castellanii* (*P≤0.01*), cytopathogenicity of THP-1 macrophages (*P≤0.05*) and pore-forming mediated cytotoxicity of sRBCs (*P≤0.05*).

3.11 Virulence assessment of *L. pneumophila* according to MLVA-8(12) genotypes and Clonal Complexes

To investigate whether the genotype affected the virulence of *L. pneumophila*, 60 environmental *L. pneumophila* isolates from the West Bank were grown to the PE phase. Five virulence assessment tests were performed; infectivity of *A. castellanii* and cytopathogenicity of THP-1 macrophages at MOI of 10, pore-forming mediated cytotoxicity at MOI of 25, heat shock of 60°C for 30min and 100mM sodium chloride sensitivity. Data shows statistically significant (ANOVA *P<0.001*) difference between the genotypes for pore forming mediated cytotoxicity **(Figure 20)**.

Infectivity of *A. castellanii*, cytopathogenicity of THP-1 macrophages, heat shock and sodium chloride sensitivity exhibit no significant difference between *L. pneumophila* genotypes. G4(17) has statistically significant (Independent t-test *P≤0.05*) in comparison with Gt10(141) and Gt40(47) for *A. castellani* infection assay.Gt64(74) is the highest infection of THP-1 macrophages (65%) has statistically significant (Independent t-test *P≤0.05*) in comparison with Gt64(74) and Gt40(47). Gt40(47) has the lowest THP-1 infection (35%) in comparison with Gt10(141) (Independent t-test *P<0.05*) **(Figure 20)**.

However, Gt4 (17) isolates showed statistically significant (Independent t-test *P≤0.001, P≤0.01 and P≤0.05*) the highest pore forming mediated cytotoxicity (60%) in comparison with Gt10 (93),Gt10(141) and Gt64(74), respectively. Moreover, Gt6 (18) isolates showed statistically significant (Independent t-test *P≤0.01*) higher than Gt10(93) and Gt10(141) (Fig.2). Furthermore, Gt40(47) showed statistically significant (Independent t-test *P≤0.01* and *P≤0.05*) higher than Gt10(93), Gt10(141) and Gt64(74) **(Figure 20)**.

Gt40(47) is sodium chloride labile genotype which has statistically significant (Independent t-test *P≤0.01* and *P≤0.05*) in comparison with Gt6(18), Gt10(93), Gt10(141) and Gt4(17) respectively. Moreover, Gt40(47) is heat labile genotype has significant difference with Gt10(141) (Independent t-test *P≤0.05*). Taken together, these results demonstrate that there is no clear significant difference between *L. pneumophila* isolates according to the MLVA-8(12) genotypes. Gt4(17) and Gt6(18) have high and constant activity **(Figure 20)**.

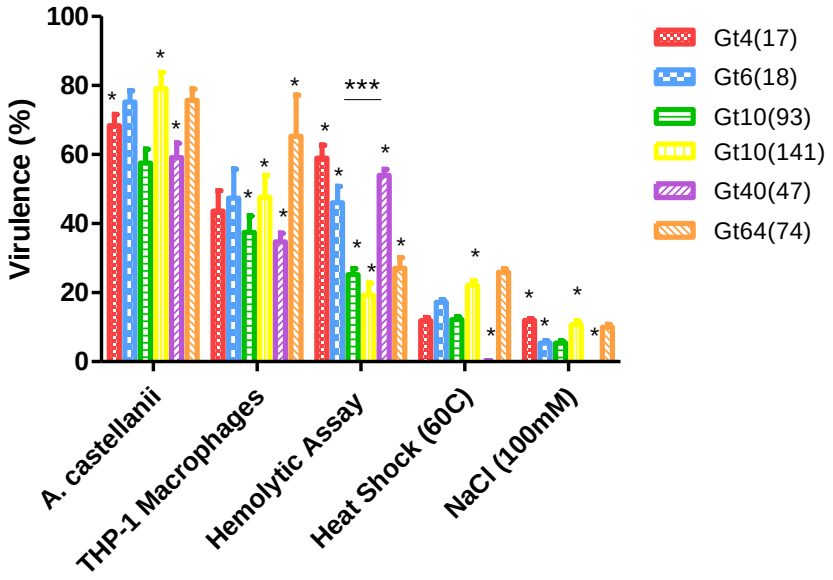


Figure 20: The virulent activity of *L. pneumophila* genotypes during post exponential phase was determined by five different virulence tests; infectivity of *A.castellanii* and THP-1 macrophages, pore forming mediated cytotoxicity of sRBC's, heat shock at 60°C for 30min and sodium chloride sensitivity. Contact-dependent hemolysis is significantly different for the genotypes. Values are means of triplicate samples, and error bars represent standard deviations. The * indicates statistically significant differences.

Regarding the clonal complexes, MLVA-8(12) genotypes in the West Bank are clustered together into 4 different clonal complexes. To investigate whether the VACCs affected the virulence of *L. pneumophila*, 60 environmental *L. pneumophila* isolates were grown at post exponential phase. The previously mentioned five virulence test were performed. Data shows statistically significant (ANOVA *P≤0.001*) difference between the four VACCs in pore forming mediated cytotoxicity test **(Figure 21).**

VACC1 isolates showed statistically significant (Independent t-test *P≤0.001*) the highest pore forming mediated cytotoxicity (53%) in comparison with VACC11 (26%). VACC5 isolates showed statistically significant (Independent t-test *P≤0.05* and *P≤0.001*) higher pore forming mediated cytotoxicity than VACC2 and VACC 11 respectively **(Figure 21).**

Moreover, VACC1 isolates showed statistically significant the highest resistant percentage for 60°C heat shock for 30min and 100mM NaCl tolerance (15% and 9% respectively) in comparison with VACC5 with (Independent t-test *P≤0.05* respectively). Also, VACC5 is labile genotype for heat and sodium tolerance (Independent t-test *P≤0.05* respectively) in comparison with the other genotypes **(Figure 21).**

In addition, VACC2 isolates showed statistically significant (Independent t-test *P≤0.05*) the highest cytopathogenecity activity of THP-1 macrophages (55%) in comparison with VACC11 (37%). Taken together, these results demonstrate when the genotypes are clustered together in clonal complexes show better significant difference between the *L. pneumophila* isolates. Also, VACC1 and VACC2 are more virulent that VACC5 and VACC11 with the exception of VACC5 in pore forming mediated cytotoxicity assay **(Figure 21).**

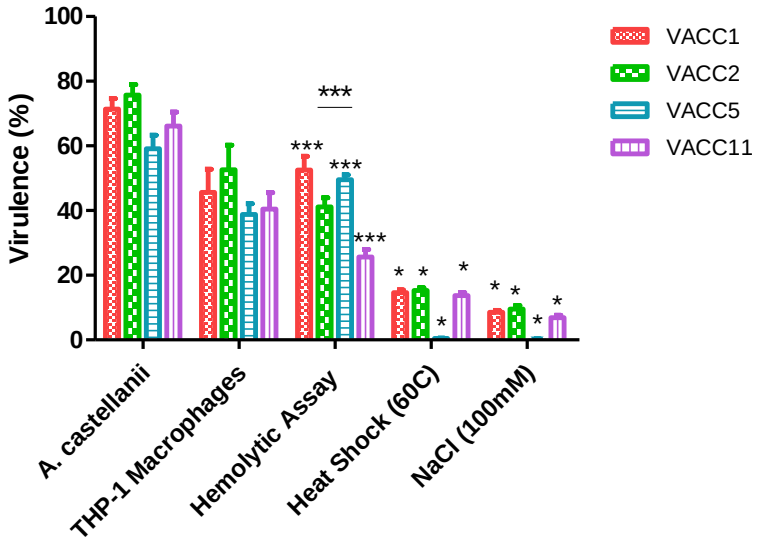


Figure 21: The virulent activity of *L. pneumophila* clonal complexes during post exponential phase was determined by five different virulence tests; infectivity of *A.castellanii* and THP-1 macrophages, pore forming mediated cytotoxicity of sRBC's, heat shock at 60°C for 30min and sodium chloride sensitivity. Virulence assessment tests are significantly different for the four clonal complexes. Values are means of triplicate samples, and error bars represent standard deviations. The * indicates statistically significant differences.

3.12 Environmental niche and virulence assessment of *L. pneumophila* isolates

To investigate whether the environmental niches affected the virulence of *L. pneumophila*, 60 environmental *L. pneumophila* isolates from the West Bank were grown to the PE phase. Five virulence assessment tests were performed; infectivity of *A. castellanii* and cytopathogenicity of THP-1 macrophages at MOI of 10, pore-forming mediated cytotoxicity at MOI of 25, heat shock of 60°C for 30min and 100mM sodium chloride sensitivity. Data shows statistically significant (ANOVA *P≤0.001*) difference between the genotypes for pore forming mediated cytotoxicity **(Figure 22)**.

Niche A isolates showed statistically significant (Independent t-test *P≤0.001*) the lowest pore forming mediated cytotoxicity (24%) in comparison with Niche B and Niche C. While the other virulence tests didn't show any difference between the niches. Taken together, environmental niches don't have a clear effect on *L. pneumophila* virulence

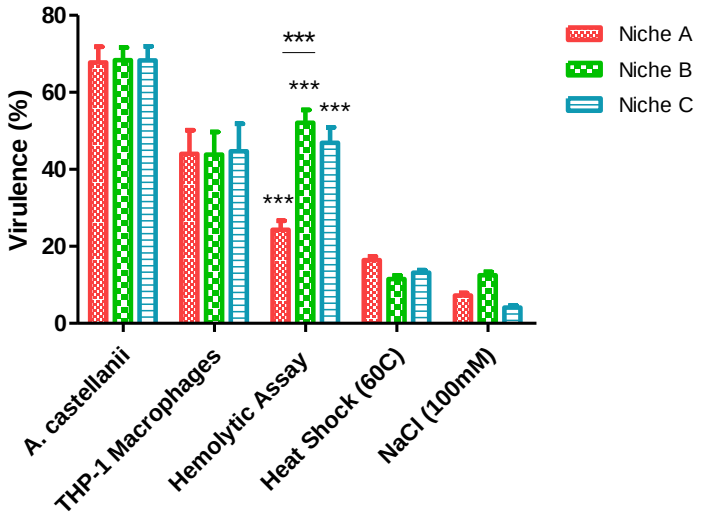


Figure 22: The virulent activity of *L. pneumophila* environmental niches during post exponential phase was determined by five different virulence tests; infectivity of *A.castellanii* and THP-1 macrophages, pore forming mediated cytotoxicity of sRBC's, heat shock at 60°C for 30min and sodium chloride sensitivity. Virulence assessment tests are significantly different for the four clonal complexes. Values are means of triplicate samples, and error bars represent standard deviations. The * indicate statistically significant differences.

3.13 General features of genomes from *L. pneumophila* isolates

Twenty-two *L. pneumophila* genomes were sequenced with approximately 100 X coverage using the Illumina Hiseq platform. A range of 28 to 103 contigs was obtained from the Hiseq runs and assembled with Velvet to reduce the gaps between the contigs. The genomes of the isolates H3_Gt14(31)_D and A194_Gt40(47)_Ps consisted of 3,691,263 bp and 3,467,907 bp, respectively, according to the PacBio sequencing

results. Strain H3 is an environmental isolate from the HZI hot water supply. It is serogroup 6 Chicago, sequence type 1431 and MLVA genotype Gt14(31). It contained genomic islands of a total length of 209,867 bp representing 5.7% of the genome. A total of 3386 genes, 3336 Coding DNA Sequence (CDS), 9 rRNA and 43 tRNA genes were identified. Strain A194_Gt40(47)_Ps is an environmental isolate from Hebron, Palestine. It is serogroup 6 Dresden and MLVA genotype Gt40(47). The total length the genomic islands of strain A194_Gt40(47)_Ps was of 156,728bp (4.5%) comprising 3121 genes, 3071 CDS, 9 rRNA and 43 tRNA genes. **Table A3** summarizes the main features of all *L. pneumophila* isolates and their genomes sequenced by the Hiseq and PacBio platform in comparison to all reference genomes of *L. pneumophila* used in this study.

3.14 SNP analysis of *L. pneumophila* isolates

A phylogenomic tree was constructed using single nucleotide polymorphism (SNP) analysis of the seventeen reference genomes available from GeneBank and our 38 *L. pneumophila* genome sequenced strains (**Table A3**). Six main branches or clusters were observed in the constructed phylogenomic tree (**Figure 23**). Branch one includes, two clinical isolates from Germany and two environmental isolates from the West Bank, Palestine. This branch included five reference *L. pneumophila* genomes (Lpn-LPE509, Lpn-Philadelphila1, Lpn-Thunderbay, Lpn-ATCC 43290 and *L. pneumophila* strain Lpn-Lpm7613). Branch two included only isolate A15_Gt12(84)_Ps and the reference strains Lpn-Lens and Lpn-Lorraine. The third branch included environmental isolates (H29_Gt22(100)_D, H34_Gt22(102)_D and H35_Gt22(102)_D) from the HZI, Germany and the New York City outbreak reference genomes (*L. pneumophila* strain Lpn-D7630, Lpn-D7631 and Lpn-D7632) plus an environmental reference strain from France (*L. pneumophila* strain Lpn-HL06041035). As expected, the vast majority of the isolates from our collection, either from the West Bank or Germany, clustered in branch four. Branch four included the worldwide distributed ST1 represented by reference strain Paris which is closely related to our MLVA genotype Gt4(17) including the strains A5_Gt4(17)_Ps and A139_Gt4(17)_Ps, Gt6(18) isolates included strains A29_Gt6(18)_Ps and A131_Gt6(18)_Ps, exclusively found in Hebron, West Bank and the environmental isolate H39_Gt4(17)_D from the HZI, Germany. In branch five a group of clinical isolates clustered together with the reference genome of strain Pontiac. Finally, the most diverse branch in our collection was branch six named Alcoy. Branch six included closely related isolates (all Gt40(47) [L04-545_Cl_Gt40(47)_D, A193_Gt40(47)_Ps, A194_Gt40(47)_Ps and A195_Gt40(47)_Ps] from the Warstein outbreak occurring in Germany during 2013. The third highly abundant isolates of MLVA genotype Gt10(93) [A108_Gt10(93)_Ps, A112_Gt10(93)_Ps, A114_Gt10(93)_Ps and A127_Gt10(93)_Ps] in our collection clustered all together. Also, group of environmental isolates from HZI genotyped Gt14(30) and Gt14(31) [H1_Gt14(31)_D, H2_Gt14(31)_D, H3_Gt14(31)_D and H23_Gt14(30)_D]. Lastly, a group of clinical isolates from Germany [L02-465_Cl_Gt27(133)_D, L06-153_Cl_Gt71(135)_D, L06-129_Cl_Gt71(135)_D, L04-041_Cl_Gt30(137)_D and L05-341_Cl_Gt8(132)_D] plus an environmental isolate [A166_Gt8(142)_Ps], from the West Bank clustered together.

This highly diverse cluster six included only two complete reference genomes (*L. pneumophila* strain Alcoy 2300/99 and Corby) (**Figure 23**).

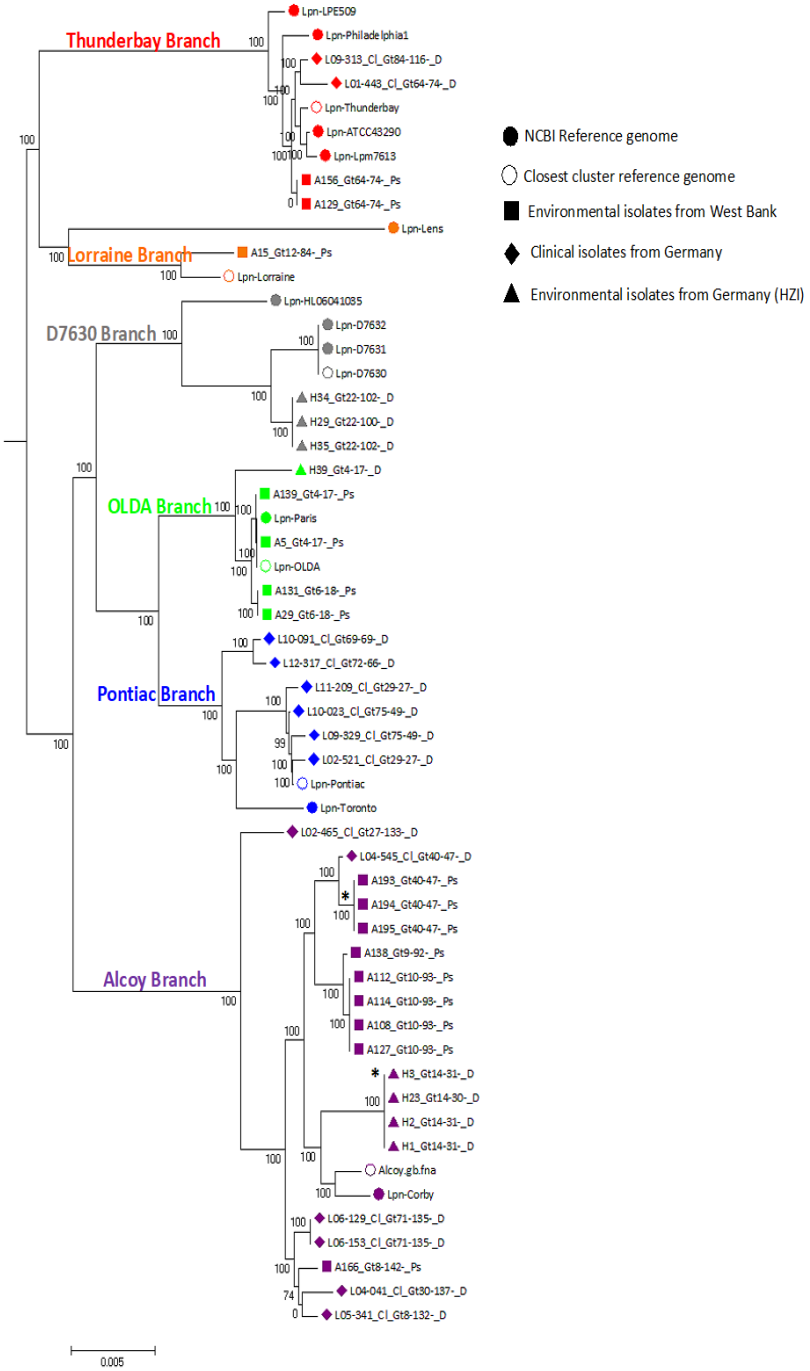


Figure 23: Phylogenomic tree based on Single Nucleotide Polymorphism (SNP) analysis of 55 *L. pneumophila* genomes. The strains group within six different clusters or branches marked in different colors. As reference for this phylogenomic tree and in order to create the core genome *L. pneumophila* strain A194_Gt40(47)_Ps was used. * indicates the two isolates (A194_Gt40(47)_Ps & H3_Gt14(31)_D) which were additionally sequenced using the PacBio platform.

Table 9: Whole-genome SNPs comparison of selected *L. pneumophila* strains in descending order.

L. pneumophila Sequenced Genomes	Core genome (%)	SNP count (bp)	Genome length (bp)
Lpn-Lens	74	3,523,725	3,345,687
Lpn-Philadelphia 1	73	2,701,132	3,397,754
Lpn-Lorraine	72	2,659,278	3,467,254
Lpn-Lpm7613	76	2,630,870	3,261,562
Lpn-Thunderbay	72	2,617,761	3,455,167
Lpn-ATCC43290	74	2,616,399	3,359,001
Lpn-LPE509	72	2,578,310	3,434,224
Lpn-HL06041035	71	2,398,444	3,492,535
Lpn-D7632	72	2,314,961	3,435,648
Lpn-D7631	72	2,314,928	3,436,178
Lpn-D7630	72	2,314,907	3,444,702
Lpn-Toronto	70	2,313,226	3,573,898
Lpn-Pontiac	70	2,223,081	3,545,001
Lpn-OLDA	71	2,177,492	3,486,108
Lpn-Paris	71	2,175,712	3,503,610
Lpn-Corby	70	1,969,316	3,576,470
H3_Gt14(31)_D	67	1,962,137	3,691,263
Lpn-Alcoy	71	1,955,470	3,516,334
A194_Gt40(47)_Ps	72	1,943,885	3,467,904

By Parsnp analysis we found 1,955,470 SNPs among the 55 strains. *L. pneumophila* strain Alcoy 2300/99 was shown to be the closest reference strain so far with a complete genome available in the GenBank database. Furthermore, the strain A194_Gt40(47)_Ps in the cluster six (named branch Alcoy according to its best reference genome) was shown to be the largest cluster in our selection consisting of altogether 21 isolates. Indeed, strain A194_Gt40(47)_Ps showed 11,585 SNPs less than performing the complete ParSNP analysis with *L.pneumophila* strain Alcoy 2300/99 as a reference. Thus, the best result for our SNP-based phylogenomic tree was retrieved with strain A194_Gt40(47)_Ps as reference and documented in **Figure 23** and **Table 9**. The greatest diversity was observed among isolates associated with cluster six, which differed by 69 to 11,755 SNPs compared to *L.pneumophila* strain A194_Gt40(47)_Ps. Hereby, one sub-lineage, consisting of isolates H1_Gt14(31)_D, H2_Gt14(31)_D, H3_Gt14(31)_D and H23_Gt14(30)_D, differed by only 5-8 SNPs compared to the strain H3_Gt14(31)_D (**Table A4**).

Within cluster one, *L. pneumophila* strain Thunderbay was most closely related to our isolates differing by 3,018 to 6,283 SNPs, respectively. In cluster two, the isolate A15 was related to *L. pneumophila* strain Lorraine and differed by 13,180 SNPs. Within cluster three, both clinical and environmental strains from the New York City outbreak showed differences between 9,764 to 9,831 SNPs. Within the worldwide distributed cluster three, *L. pneumophila* strain OLDA was the most closely related reference genome to our isolates of MLVA genotype Gt4(17) and Gt6(18) differing by 76 to 11,020 SNPs. Lastly, within cluster five *L. pneumophila* strain Pontiac had the most closely related reference genome differing by 823 to 14,949 SNPs to our isolates (**Table A4**).

In a detailed bioinformatics analysis, cluster one contained four isolates, three of which belong to Gt64(74) and the clinical isolate from Germany L09-313_Cl_Gt84(116)_D was Gt84(16). The two environmental isolates from the West Bank [A156_Gt64(74)_Ps and A129_Gt64(74)_Ps] were close to identical in terms of SNPs (up to 4 SNPs). However, the clinical isolate L09-313_Cl_Gt84(116)_D was highly diverse comprising 6,283 SNPs. Isolates of the MLVA genotypes Gt22(100) and Gt22(102) formed a specific sub-lineage but could be still distinguished in most cases. For instance, SNP analysis revealed that some Gt22(100) and Gt22(102) isolates differed by 67 SNPs. On the other hand, in cluster four sub-lineage the genome sequence of isolate A29_Gt6(18)_Ps differed by only 2 SNPs from the environmental isolate A131_Gt6(18)_Ps. Interestingly, isolates A29_Gt6(18)_Ps and A131_Gt6(18)_Ps were isolated from the hospital G located in Hebron in the Southern West Bank (**Table A4**).

The MLVA genotype Gt4(17) formed a specific sub-lineage in cluster four, but isolates of Gt4(17) from different cities could still be distinguished because they have been isolated in different years. For example, the strains A5_Gt4(17)_Ps and A139_Gt4(17)_Ps differed by 119 SNPs. However, the strain H39_Gt4(17)_D, also being Gt4(17), an environmental isolate from HZI, Germany showed a highly different SNPs pattern of 11,020 SNPs. The branch five contained six clinical isolates from Germany. Two of which were Gt75(49), another two were Gt29(27), one was Gt72(66) and the last one was Gt69(69). These branch five isolates showed a highly different SNP pattern ranging from 823 to 14,949 SNPs among different clinical strains across Germany (**Table A4)**.

In branch six, isolates of Gt40(47) formed MLVA-genotype specific sub-lineages, but closely related outbreak associated isolates from Warstein could still be distinguished in most cases. For example, SNP analysis revealed that isolates of Gt40(47) associated with the Warstein outbreak differed by 595 SNPs. Peltzold *et al* showed SNPs analysis of targeted genes revealed an average of two SNPs per 100bp (131). In the isolates from Palestine in hospital H a higher SNP diversity (2,822 SNPs) was identified in comparison to the clinical isolate L04-545_Cl_Gt40(47)_D. Similarly, isolates of Gt10(93) clustered in one sub-lineage but clearly differed by 541 SNPs. Furthermore, the genome sequence of isolate L06-153_Cl_Gt71(135)_D which was Gt71(135) differed by only 3 SNPs from isolate L06-129_Cl_Gt71(135)_D (**Figure 23** and **Table A4**).

There is a good concordance between phylogenomic clusters and MLVA-genotypes. Cluster one corresponds to Gt64(74). Cluster three included the MLVA genotypes Gt22(100) and Gt22(102) which were isolated from the HZI drinking water supply. The worldwide present and highly virulent cluster four was inclusive for Gt4(17) and Gt6(18). Cluster five corresponds to Gt29(27), Gt69(69), Gt72(66) and Gt75(49). Lastly, the highly diverse cluster six included Gt40(47), Gt9(92), Gt10(93), Gt14(40), Gt14(31), Gt71(135), Gt8(132), Gt8(142) and Gt30(137) (**Table A3**). However, genome sequencing could distinguish all isolates in this cluster. For instance, SNP analysis differentiated between every *L. pneumophila* strain in the same phylogenomic branch.

An alternative way to evaluate variation among the genomes is to study single base polymorphism in detail. Therefore, we visualized Parsnp results with the software Gingr. High quality SNPs extracted from those positions having SNP changes in the *L. pneumophila* targeted genes. There were four isolates of Gt14(30) and Gt14(31) namely H1_Gt14(31)_D, H2_Gt14(31)_D, H3_Gt14(31)_D and H23_Gt14(30)_D, all isolated from the hot water supply of the HZI, some isolates had been isolated in different years, i. e. 2009 and 2013. These isolates differed by just nine SNPs. We identified each individual SNP among these strains (**Table 10**).

Table 10: Nucleotide Polymorphism differences in cluster six sub-branch of *L. pneumophila* isolates from the HZI cooling tower relative to the H3_Gt14(31)_D-PacBio reference sequence.

Nucleotide Polymorphism	**Label (locus tag)**	***L. pneumophila* Isolate**	**Gene Name**	**Gene Product**	**Function**	**Length (bp)**	**Virulence Similarity (Id)**
A → T	H3_00006	H2_Gt14(31)_D	SpeA	Biosynthetic arginine decarboxylase	Intracellular Growth	1,891	NV
C → T	H3_00027	H23_Gt14(30)_D	YhdG_1	Putative amino acid permease	Survival *in vivo*	1,391	NV
C → T	H3_00935	H1_Gt14(31)_D, H2_Gt14(31)_D and H23_Gt14(30)_D	BepC	Outer membrane efflux protein	Survive inside the host cells, restores drug efflux	1,724	NV
T → G	H3_00978	H1_Gt14(31)_D	mmgC	Acyl-CoA dehydrogenase	0	1,154	NV
A → G	H3_01656	H23_Gt14(30)_D	0	Putative hydrolase	0	983	NV
C → T	H3_02694	H1_Gt14(31)_D, H2_Gt14(31)_D and H23_Gt14(30)_D	(ankF/legA14/ceg31)	Dot/Icm type IV secretion system effector (Ankyrin repeats)	Host microbes interactions, Protein-Protein interaction, Intracellular replication	2,765	96
G → T	H3_02706	H23_Gt14(30)_D	0	Aspartyl/Asparaginyl beta hydrolase	0	719	NV
C → T	H3_02929	H1_Gt14(31)_D, H2_Gt14(31)_D and H23_Gt14(30)_D	0	Phagosome trafficking protein DotA	Bacterial replication, virulence, Intracellular growth, Evasion of the Endocytic pathway	3,119	79
C → T	H3_03123	H23_Gt14(30)_D	0	Hypothetical protein	0	1,289	NV

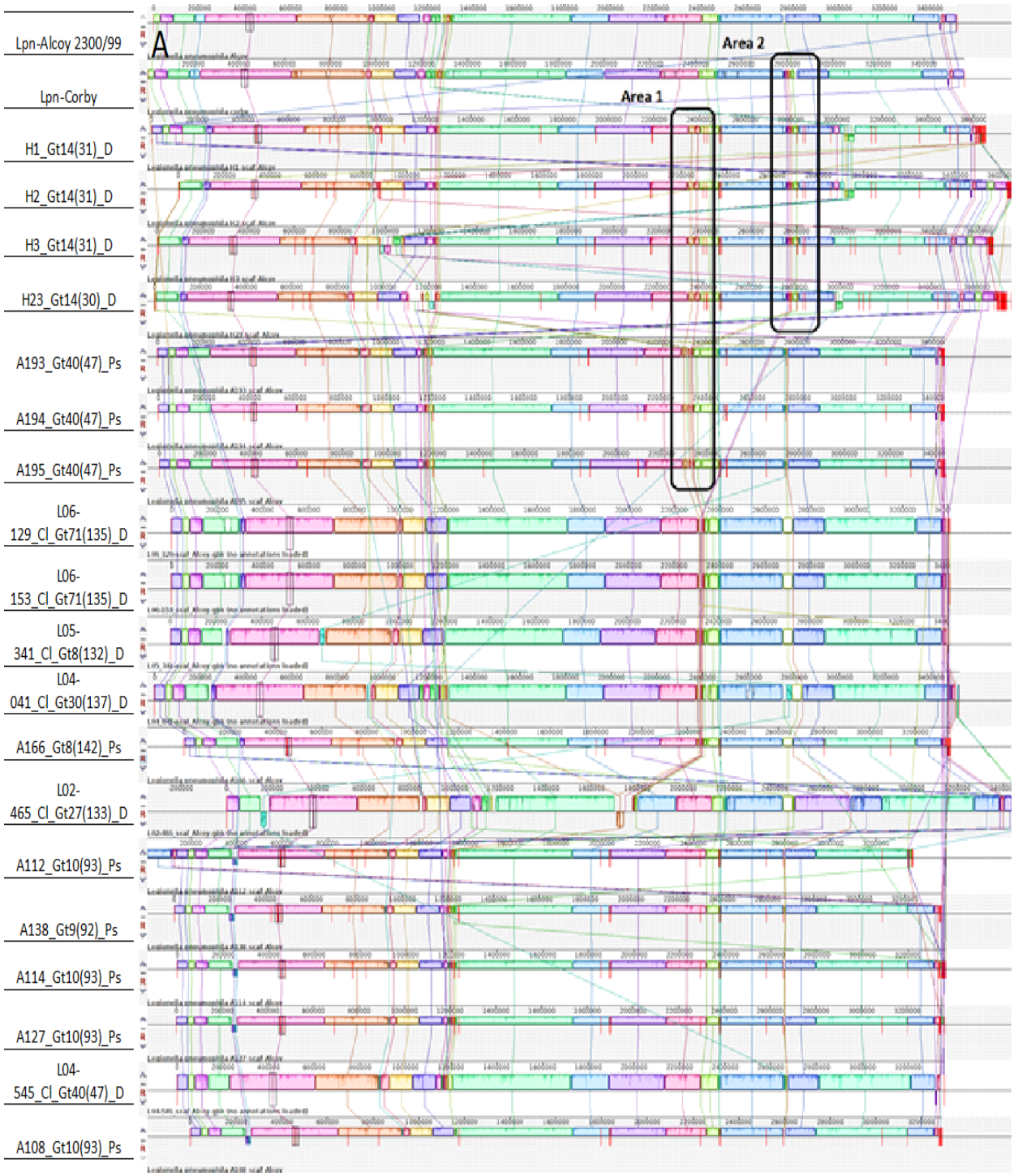
NV : Non-Virulence Gene

A detailed bioinformatics comparison of nine genes revealed substantial single nucleotide polymorphism if the genome of isolate H3 was used a reference genome (H3_00006, H3_00027, H3_00935, H3_00978, H3_01656, H3_02694, H3_02706, H3_02929 and H3_03123) **(Table 10**). These nine genes were blasted (Blastp) against *L. pneumophila* strain Philadelphia1 as default reference genome in the Virulence Factor Data Base (VFDB) webpage to detect virulence genes. Two virulent genes (H3_02694 and H3_02929) were identified with 96% and 79% identity by comparison with the Dot/Icm type IV secretion system effector (ankyrin repeats) and the phagosome trafficking protein dotA, respectively. These results showed that *L. pneumophila* contained different levels of diversity to be observed in virulence and other genes. There are several evolutionary scenarios that will be dealt with in the discussion to explain these observed differences.

3.15 Identification of mobile elements in *L. pneumophila* Alcoy branch

A multiple whole genome alignment using Mauve revealed largely syntenic Locally Collinear Blocks (LCB) across twenty-one genomes in cluster six, i. e. the Alcoy branch (**Figure 24**). Notably, this cluster contained clinical and environmental strains from two geographical regions (West Bank and Germany) that were highly similar except for two interesting areas. To investigate the genomic diversity of cluster six, we used the Illumina Hiseq genome sequences of all twenty-one isolates and compared them with the complete genomes of the reference strains Alcoy 2300/99 and Corby plus the two genome sequences of isolate H3_Gt14(31)_D and A194_Gt40(47)_Ps obtained by the PacBio platform.

Genome size and all main genome features between these strains are summarized in (**Table A3**). In this study, we focused on two different areas that were clearly identified in H3_Gt14(31)_D, A194_Gt40(47)_Ps and Lpn-Corby strains. Area one was a 84 kbp region containing regulatory proteins, antibiotic resistance proteins, genomic islands and three tRNAs (**Figure 24** and **Table 11**). This area does not exist in the reference genomes of Alcoy 2300/99 and Corby but was clearly present in the isolates H1_Gt14(31)_D, H2_Gt14(31)_D, H3_Gt14(31)_D, H23_Gt14(30)_D, A193_Gt40(47)_Ps, A194_Gt40(47)_Ps and A195_Gt40(47)_Ps. Also it was present partially in the isolates L04-041_Cl_Gt30(137)_D, L05-341_Cl_Gt8(132)_D, L06-129_Cl_Gt71(135)_D, L06-153_Cl_Gt71(135)_D and A166_Gt8(142)_Ps. The second area is with 49 kbp much smaller was present in the Corby genome and the isolates H1_Gt14(31)_D, H2_Gt14(31)_D, H3_Gt14(31)_D and H23_Gt14(30)_D (**Figure 24**). All major gene components of area two are summarized in **Table 11**. Altogether the twenty-one isolates in cluster six are highly syntenic. There are few interesting exceptions which were identified as mobile elements.


A
Area 1
Area 2
Lpn-Alcoy 2300/99
Lpn-Corby
H1_Gt14(31)_D
H2_Gt14(31)_D
H3_Gt14(31)_D
H23_Gt14(30)_D
A193_Gt40(47)_Ps
A194_Gt40(47)_Ps
A195_Gt40(47)_Ps
L06-129_Cl_Gt71(135)_D
L06-153_Cl_Gt71(135)_D
L05-341_Cl_Gt8(132)_D
L04-041_Cl_Gt30(137)_D
A166_Gt8(142)_Ps
L02-465_Cl_Gt27(133)_D
A112_Gt10(93)_Ps
A138_Gt9(92)_Ps
A114_Gt10(93)_Ps
A127_Gt10(93)_Ps
L04-545_Cl_Gt40(47)_D
A108_Gt10(93)_Ps

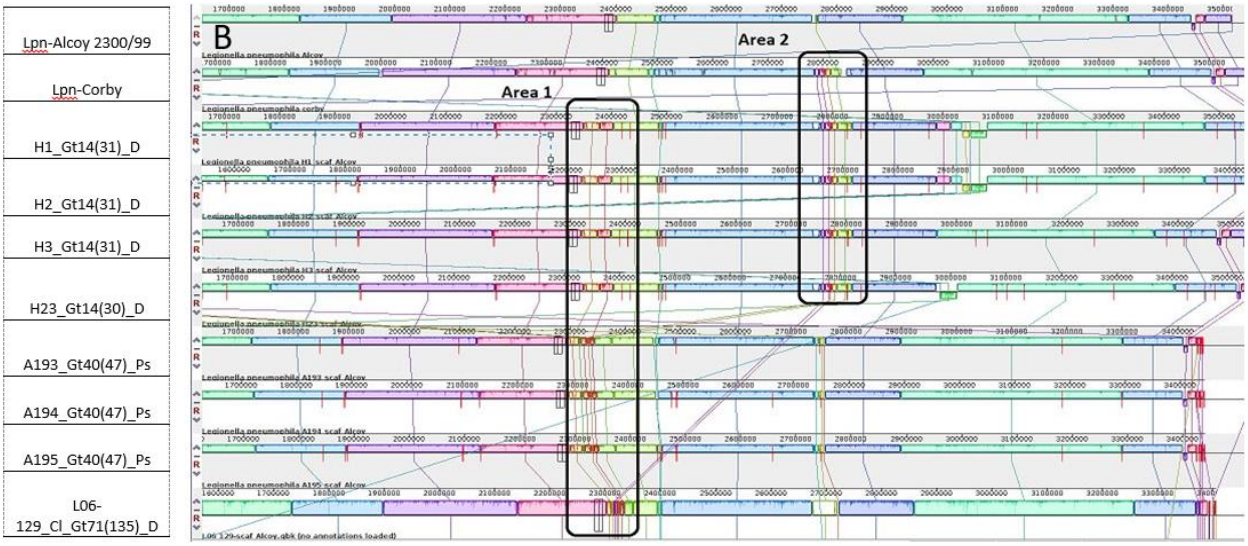


Figure 24: Mauve whole-genome alignment of *L. pneumophila* strains within cluster six. A) Overview, a Progressive Mauve was used to compare the complete genomes of the strains Alcoy 2300/99 and Corby as well as cluster six isolates. Shown with shared common colors across genomes are pairwise LCBs (locally collinear blocks). The program was run using default parameters as described in (122). The cluster six organization, as well as the identity and location of the ~84-kb area 1 and the ~49-kb area 2 elements are shown. **B)** Magnified in Mauve.

Table 11: Gene content according to MAUVE aligner in the studied areas. Hypothetical proteins were excluded

Gene Name	Product	Label	Length (bp)
Area 1			
	Acetyltransferase (GNAT) family protein	A194_02079	437
dpnM	Modification methylase DpnIIA	A194_02080	1,412
dlpA_1	DlpA protein	A194_02081	719
	Phosphotransferase enzyme family protein	A194_02082	968
	aminoglycoside phosphotransferase	A194_02083	1,436
	shikimate kinase	A194_02085	779
	Enterochelin esterase	A194_02086	1,415
abcT3_4	ABC transporter, ATP binding protein	A194_02088	1,547
	transcriptional regulator, MerR family, mercury resistance	A194_02090	944
	beta-lactamase AmpS	A194_02091	800
ftsI4_2	cell division protein FtsI/penicillin binding protein 2	A194_02092	1,922
mecI_2	Methicillin resistance regulatory protein mecI	A194_02093	422
	transcriptional regulator SkgA, mercury resistance	A194_02094	749
	acetyltransferase, GNAT family	A194_02097	602
	reverse transcriptase	A194_02098	1,556
	acetyltransferase, GNAT family	A194_02100	482

grpB_2	glutamate rich protein GrpB	A194_02101	956
	nucleotidyltransferase PLUS glutamate rich protein GrpB PLUS ribosomal protein alanine acetyltransferase	A194_02102	794
	pyridoxamine 5'-phosphate oxidase	A194_02104	581
	phage repressor	A194_02105	622
lvrA_2	Legionella vir region protein	A194_02106	872
lvrC	Legionella vir region protein	A194_02107	197
	putative exported protein	A194_02108	434
	Type IV secretory protein VirB4 component	A194_02121	2,774
	TraU protein	A194_02122	995
	membrane protein, Tfp pilus assembly, pilus retraction ATPase PilT	A194_02123	1,382
	conjugative coupling factor TraD	A194_02127	2,006
	Putative HTH-type transcriptional regulator/MT0914	A194_02129	572
	Avidin family protein	A194_02130	413
	putative secreted esterase	A194_02132	1,556
	Bacterial regulatory proteins, luxR family	A194_02133	800
murE3_1	UDP-N-acetylmuramyl tripeptide synthase	A194_02134	1,226
aph_2	spectinomycin phosphotransferase	A194_02137	995
	glyoxylase domain hypothetical protein	A194_02138	365
	short chain dehydrogenase	A194_02139	734
yafP	putative N-acetyltransferase YafP	A194_02140	1,697
fni	isopentenyl-diphosphate delta-isomerase	H3_02136	1,028
mvaA	hydroxymethylglutaryl CoA reductase	H3_02137	1,298
	tRNA-Lys	H3_02139	75
	tRNA-Lys	H3_02140	75
	tRNA-Arg	H3_02141	76
	site specific recombinase	H3_02142	1,160
nucS	Endonuclease NucS	H3_02143	1,004
	Acetyltransferase (GNAT) family protein	H3_02144	485
	proline/betaine transport protein like protein	H3_02145	1,274
	putative acetyltransferase	H3_02146	512
	lipolytic enzyme	H3_02147	686
	transcription regulator protein, response regulator containing CheY-like receiver domain and HTH DNA-binding domain	H3_02148	803
tylM1	dTDP-3-amino-3,6-dideoxy-alpha-D-glucopyranose N,N-dimethyltransferase	H3_02150	824
	topology modulation protein	H3_02152	542
	aminoglycoside 6'-N-acetyltransferase	H3_02154	545
	acetyltransferase, GNAT family	H3_02155	554

	transcriptional regulator, MerR family, mercury resistance	H3_02156	1,040
	methyltransferase	H3_02157	881
	Acetyltransferase (GNAT) family protein	H3_02158	419
	recombination protein F	H3_02159	1,160
	Cupin domain protein	H3_02160	386
	O-antigen acetylase	H3_02162	1,148
	serine/threonine protein kinase	H3_02163	1,001
phaB_4	acetyoacetyl CoA reductase	H3_02164	782
	sepiapterin reductase	H3_02165	794
dpnM	Modification methylase DpnIIA	H3_02165	1,412
dlpA_1	DlpA protein	H3_02168	716
	Phosphotransferase enzyme family protein	H3_02169	968
	aminoglycoside phosphotransferase	H3_02170	1,436
aroK_2	shikimate kinase	H3_02172	779
mdtK	Multidrug resistance protein MdtK	H3_02173	1,364
	transcriptional regulator, LysR family	H3_02174	863
	major facilitator superfamily transporter	H3_02175	1,175
	Peptidase family S41	H3_02176	1,337
	phage repressor	H3_02177	662
lvrA_2	Legionella vir region protein	H3_02178	872
lvrC_2	Legionella vir region protein	H3_02179	197
	putative exported protein	H3_02180	434
	Type IV secretory protein VirB4 component	H3_02193	2,774
	TraU protein	H3_02194	977
	membrane protein, Tfp pilus assembly, pilus retraction ATPase PilT	H3_02195	1,382
	membrane protein	H3_02196	1,568

Area 2			
	lipolytic enzyme	corby_02494	683
	transcription regulator protein, response regulator containing CheY-like receiver domain and HTH DNA-binding domain	corby_02495	803
cadA_2	cadmium translocating P-type ATPase CadA	Corby_02499	503
	cadmium efflux ATPase	Corby_02500	1,433
cadA_3	cadmium translocating P-type ATPase CadA	Corby_02502	2,135
helA_2	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	Corby_02503	3,149
helB_2	cation efflux system HelB	Corby_02504	1,256
helC_2	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	Corby_02505	1,244
	reverse transcriptase	Corby_02506	1,364
	reverse transcriptase	Corby_02511	1,199
	transposase IS4	Corby_02512	1,460
	phage repressor	Corby_02513	662
lvrA_2	Legionella vir region protein	Corby_02514	872
lvrB_2	Legionella vir region protein	Corby_02515	383
lvrC	Legionella vir region protein	Corby_02516	197
	putative exported protein	Corby_02517	434
	exported membrane protein	Corby_02528	1,313
	Type IV secretory protein VirB4 component	Corby_02530	2,774
	TraU protein	Corby_02531	995
	membrane protein, Tfp pilus assembly, pilus retraction ATPase PilT	Corby_02532	1,382
	membrane protein	Corby_02533	1,568
	conjugative coupling factor TraD	Corby_02536	2,006

3.16 Identification and comparison of genomic islands in *L. pneumophila* isolates

The different bioinformatics tools for identification of horizontally transferred genes, such as IslandPath-DIMOB, SIGI-HMM and IslandPick, are integrated by the IslandViewer4 webpage into one output interface. For integrated genomic island output, we combined different approaches of IslandViewer4 to identify the major GIs in the *L. pneumophila* isolates.

In cluster one, the isolates A129 and A156 were aligned against *L. pneumophila* subsp. *pneumophila* ATCC 43290, due to the fact that the closest isolate reference genome *L. pneumophila* strain Thunderbay does not exist in the IslandViewer4 webpage database. This analysis revealed four genomic islands which are not present in isolate A156 (**Figure 25**). Despite, the fact that strains A129 and A156 were in the same sub-lineage of cluster one (**Figure 23**) and display a rather similar number of SNPs against most other reference genomes in this cluster (**Table A4**). Isolate A129 contained a

total of 294 kbp genomic islands (8.3% of the whole chromosome). However, isolate A156 only contained genomic islands of a size of 174 kbp (5.2%). Isolate A129 contained the whole genomic content of isolate A156 plus extra-genetic islands containing around 40% hypothetical proteins (**Table A7**). This extra-genetic area in isolate A129 could be located extra-chromosomally in a plasmid.

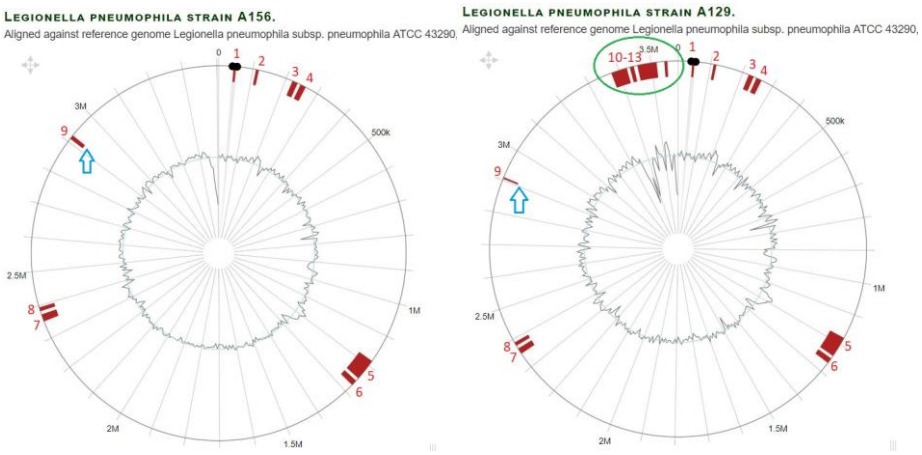


Figure 25: Comparison of Genomic islands in cluster 1 of *L. pneumophila* strain A129 and A156 aligned against reference genome *L. pneumophila* subsp. *pneumophila* ATCC 43290. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii)Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow). For a summary see **Table A5**.

In cluster three, the three environmental isolates from HZI contained 8-10 genomic islands (**Figure 26**). Isolate H35 contained two extra-genomic islands (GI4 and GI10). The three isolates shared more than 95% of their genomic island genes (**Table A8**). **Table A5** summarizes the number of genomic islands and genomic island genes for all isolates.

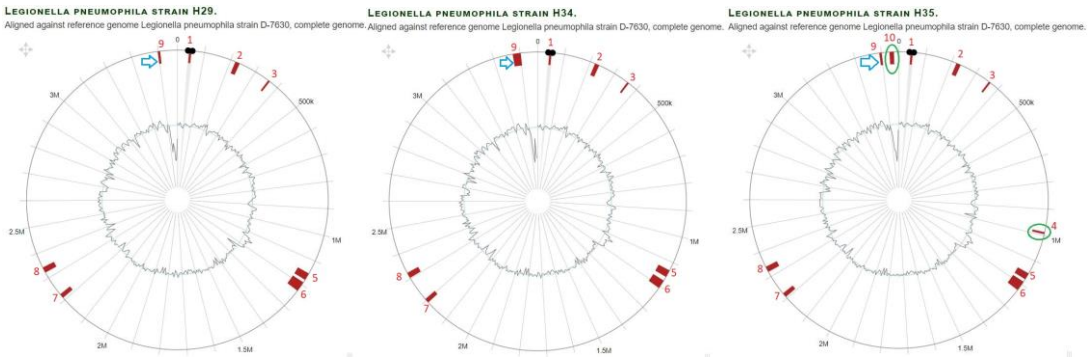


Figure 26: Comparison of Genomic islands in cluster 3 of *L. pneumophila* strain H29, H34 and H35 aligned against reference genome *L. pneumophila* strain D-7630. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii)Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow). For a summary see **Table A5**.

In cluster four, isolate A139 contained eleven genomic islands whereas isolate A5 contained ten genomic islands (**Figure 27**). However, GI11 in isolate A5 was split into GI1 and GI11 in isolate A139. Both isolates shared the same genomic islands of a

length 226 kbp (6.3%) and 224 kbp (6.3%), respectively (**Table A3**). On the contrary, isolate A29 isolate contained twelve genomic islands and isolate A131 contained thirteen genomic islands (**Table A5** and **Figure 28**). Despite the fact that the isolates A29 and A131 were in the same sub-lineage, only 2 SNPs were different against the reference strains OLDA and Paris (**Table A4**). Isolate A29 contained 339 kbp (9.3%) of genomic islands in contrast to isolate A131 containing 281 kbp (7.9%) of genomic islands (**Table A3**). In cluster five, the isolates L11-209, L09-329, L02-521, L12-317 and L10-091 contained 6, 7, 8, 10 and 13 genomic islands respectively (**Table A5**). The clinical isolate L12-317 contained 187 kb (5.3%) genomic islands and the isolates L09-329 and L02-521 contained 117kb (3.4%) and 118kb (3.4%), respectively (**Table A3**).

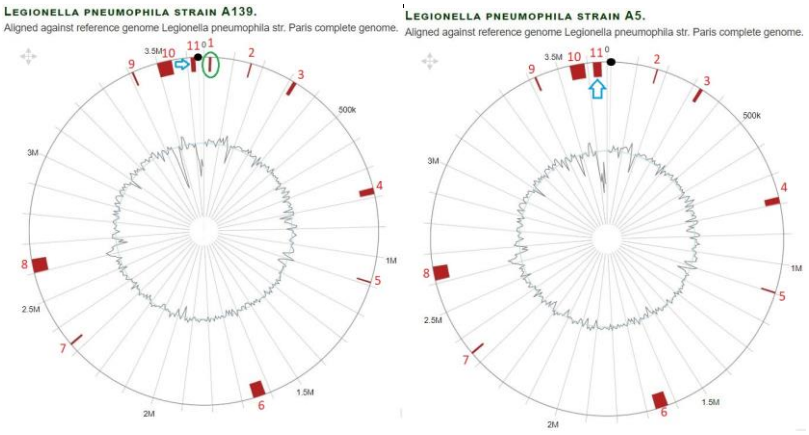


Figure 27: Comparison of Genomic islands in cluster 4 of *L. pneumophila* strain A5 and A139 aligned against reference genome *L. pneumophila* strain Paris. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii)Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow). For a summary see **Table A5**.

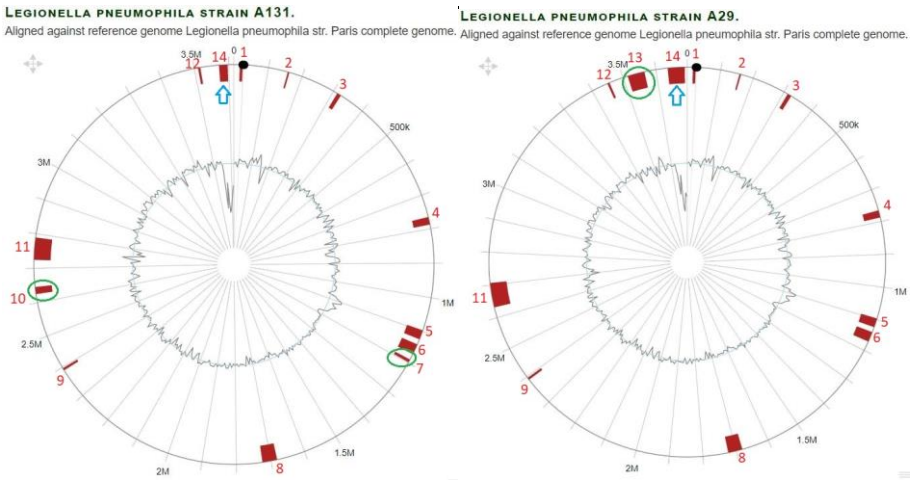


Figure 28: Comparison of Genomic islands in cluster 4 of *L. pneumophila* strain A29 and A131 aligned against reference genome *L. pneumophila* strain Paris. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii)Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow). For a summary see **Table A5**.

In cluster six, the isolates L04-545, A193 and A195 contained eight genomic islands (**Figure 29**). However, isolate A194 contained six genomic islands (**Table A5**). GI1 and GI4 did not exist in isolate A194. GI3 translocated as shown in (**Figure 29**). Interestingly, isolates of Gt40(47) from the West Bank and Germany were closely related to isolates from the Warstein outbreak showed a high similarity in the GI gene content (**Table A9**), genomic islands and SNPs pattern which differed by less than 1500 SNPs if aligned against reference genomes of strain Alcoy 2300/99 and Corby (**Table A4**). These results demonstrated a high similarity between environmental isolates from Hebron (West Bank) and the clinical isolate from Germany. Moreover, the sub-lineage comprising the isolates A108 and A138 contained seven genomic islands. GI4 was not present in isolates A112, A114 and A127 (**Figure 30**). Isolate A138 contained 163 kbp (4.8%) of genomic island genes and isolate A127 contained 143kbp (4.3%) of genomic island genes.

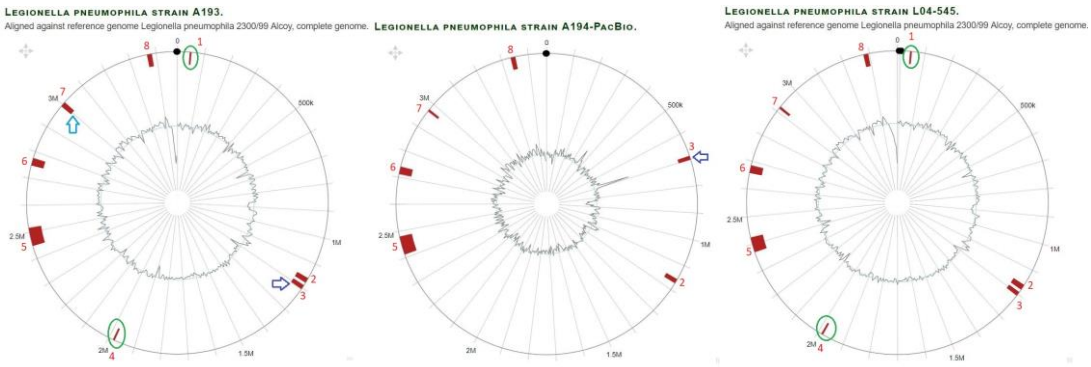


Figure 29: Comparison of Genomic islands in cluster 6 of *L. pneumophila* strain A193, A194 and L04-545 (also represents A195 not shown) aligned against reference genome *L. pneumophila* 2300/99 Alcoy. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii) Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow); (v) GI translocation (purple arrow). For a summary see **Table A5**.

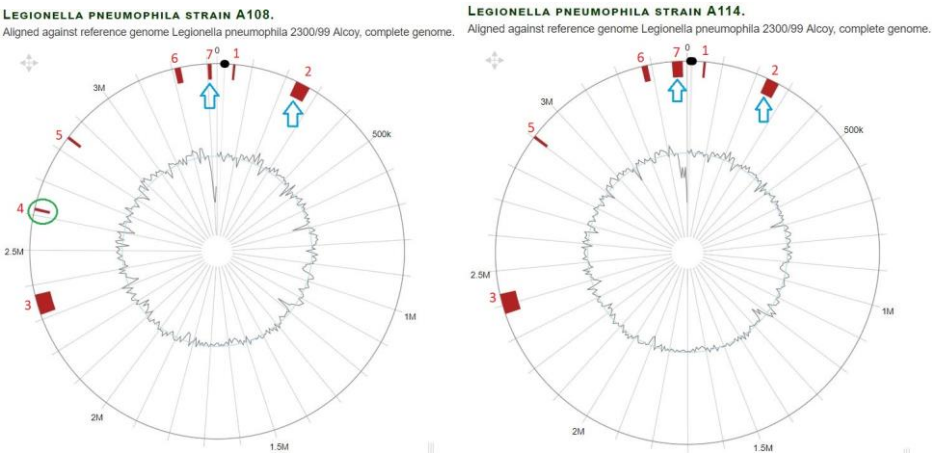


Figure 30: Genomic islands in cluster 6 branch *L. pneumophila* A108 (also represents A138 not shown) and A114 (also representing A112 and A127 not shown) aligned against reference genome *L. pneumophila* 2300/99 Alcoy. The outer layers, the circle shows: (i) nucleotide positions in megabases (M) (black); (ii)Island Viewer-annotated potential genomic islands (GI) are labeled accordingly (red); (iii) Extra GIs (marked Green oval); (iv) GI different size (Blue arrow). For a summary see **Table A5**.

The sub-lineage of the HZI isolates in cluster six showed many differences in GIs. Isolate H1 contained ten genomic islands and 253kb (6.9%) genomic island genes while isolate H23 contained thirteen genomic islands and 373kb (9.9%) genomic island genes (**Table A5**). Hypothetical proteins formed the major part of the island genes, especially in isolate H23 they comprised 45% of the gene sequences. The isolates from the hot water supply of the HZI differed by less than 10 SNPs (**Table A4**), had the same MLVA genotype and clustered closely together in the phylogenomic tree (**Figure 23** and **Table A3**). However, the HZI isolate sub-lineage has major differences in genomic islands. Finally, clinical isolates in the cluster six sub-lineage contained a variety of genomic islands and genomic island genes. For example, isolate L06-129 contained only five genomic islands while isolate L04-041 contained 14 genomic islands. Hypothetical proteins formed the major part ofGI genes for all these isolates (**Table A5**).

L. pneumophila isolates revealed high genome plasticity by gaining genes from genomic islands (GI). Three different isolates A194_Gt40(47)_Ps, H35_Gt22(102)_D and L02-465_Cl_Gt27(133)_D gained genes contributing to adaptation and pathogenicity containing heme proteins, cytochrome C proteins, and hemin binding protein. Isolate A194_Gt40(47)_Ps gained 15kb bifunctional hemolysin gene. Clinical isolates L10-091_Cl_Gt69(69)_D and L06-129_Cl_Gt71(135)_D gained heat shock proteins helping *L. pneumophila* to improve its virulence by adapting to high body temperature. Also, L06-129_Cl_Gt71(135)_D gained CRISPR-Cas helping the strain to improve its defense mechanisms. Isolate A29_Gt6(18)_Ps contained cold shock protein helping it to adapt to cold environments. Strain A112_Gt10(93)_Ps gained anaerobic sulfatase-maturating enzyme helpings to adapt to the environment. The Gt14(30) and Gt14(31) environmental isolates from HZI sub-lineage gained beta-lactamase making them antibiotic resistant The environmental isolates from the West Bank A5_Gt4(17)_Ps and A139_Gt4(17)_Ps and environmental isolate from HZI

H1_Gt14(31)_D gained bacteriophage helping *L. pneumophila* to increase pathogenicity and adapt faster in the environment to changing conditions **(Table 12).**

Table 12: Unique (interesting) genes in *L. pneumophila* genomic islands as identified by IslandViewer4

Cluster No.-Reference genome	*L. pneumophila* Isolate	Gene ID	Locus tag	Gene Length (bp)	Product
Branch 3-D7630	H34_Gt22(102)_D	0	H34_02359	965	major outer membrane protein
		tufA_2	H34_02301	917	translation elongation factor Tu (EF-Tu)
		0	H34_02302	341	preprotein translocase subunit SecE
		nusG	H34_02303	548	transcription antitermination protein NusG
		rplK	H34_02304	434	50S ribosomal protein L11
		rplA	H34_02305	695	50S ribosomal protein L1
		rplJ	H34_02306	533	50S ribosomal protein L10
		rplL	H34_02307	380	50S ribosomal protein L7/L12
		rpoB	H34_02308	4,106	DNA-directed RNA polymerase beta subunit
		rpoC	H34_02309	4,247	DNA-directed RNA polymerase subunit beta'
		rpsL	H34_02310	380	30S ribosomal protein S12
		rpsG	H34_02311	527	30S ribosomal protein S7
		fusA	H34_02312	2,084	translation elongation factor G (EF-G)
	H35_Gt22(102)_D	ccmC	H35_00671	755	heme exporter protein CcmC
		ccmD	H35_00670	158	heme exporter protein CcmD
		ccmE	H35_00669	431	cytochrome c-type biogenesis protein CcmE
		ccmF	H35_00668	1,952	cytochrome c-type biogenesis protein CcmF
		ccmG	H35_00667	533	cytochrome C biogenesis protein
		ccmH	H35_00666	401	c-type cytochrome biogenesis protein CcmH
		cycH	H35_00665	686	cytochrome c type biogenesis protein CcmH
Branch 4-Paris	A29_Gt6(18)_Ps	0	A29_00568	206	cold shock domain family protein
	A5_Gt4(17)_Ps	0	A5_02183	986	Microvirus H protein (pilot protein)
		0	A5_02184	1,541	Bacteriophage replication gene A protein (GPA)
		0	A5_02185	260	Phage protein C
		0	A5_02186	458	Bacteriophage scaffolding protein D
		0	A5_02187	116	Microvirus J protein
	A139_Gt4(17)_Ps	0	A139_02127	1,025	Bacteriophage replication gene A protein (GPA)
		0	A139_02128	260	Phage protein C
		0	A139_02129	458	Bacteriophage scaffolding protein D
		0	A139_02130	116	Microvirus J protein
		0	A139_02131	1,283	Capsid protein (F protein)
		0	A139_02132	527	Major spike protein (G protein)
		0	A139_02133	986	Microvirus H protein (pilot protein)

Branch 5-Pontiac	L10-091_Cl_Gt69(69)_D	hspC2_1	L10-091_02122	491	small heat shock protein HspC2
		0	L10-091_02123	587	heat shock hsp20
Branch 6-Alcoy	L02-465_Cl_Gt27(133)_D	0	L02-465_00734	425	hemin binding protein Hbp
	A194_Gt40(47)_Ps	cya	A194_00644	15,197	Bifunctional hemolysin/adenylate cyclase precursor
	A112_Gt10(93)_Ps	chuR	A112_02425	1,475	Anaerobic sulfatase-maturating enzyme
	H23_Gt14(30)_D	ampG	H23_01834	1,247	beta lactamase induction signal transducer AmpG
		0	H23_02025	1,361	metallo-beta lactamase family protein
		0	H23_01996	1,424	Bacteriophage replication gene A protein (GPA)
		0	H23_01997	260	Phage protein C
		0	H23_01998	458	Bacteriophage scaffolding protein D
		0	H23_01999	116	Microvirus J protein
		0	H23_02000	1,283	Capsid protein (F protein)
		0	H23_02001	527	Major spike protein (G protein)
		0	H23_02002	986	Microvirus H protein (pilot protein)
	H1_Gt14(31)_D	ampG	H1_02766	1,247	beta lactamase induction signal transducer AmpG
		0	H1_00952	1,361	metallo-beta lactamase family protein
		0	H1_02730	416	Bacteriophage scaffolding protein D
		0	H1_02731	116	Microvirus J protein
		0	H1_02732	1,283	Capsid protein (F protein)
		0	H1_02733	527	Major spike protein (G protein)
		0	H1_02734	986	Microvirus H protein (pilot protein)
		0	H1_02735	1,541	Bacteriophage replication gene A protein (GPA)
	H2_Gt14(31)_D	0	H2_02312	1,361	metallo-beta lactamase family protein
	H3_Gt14(31)_D	0	H3_03288	1,361	metallo-beta lactamase family protein
	L06-129_Cl_Gt71(135)_D	htpX_3	L06-129_01854	965	heat shock protein, protease HtpX
		0	L06-129_01853	470	Phosphate-starvation-inducible E
		trxA3_2	L06-129_01852	440	thioredoxin
		0	L06-129_01848	569	small HspC2 heat shock protein
		cas1	L06-129_01845	968	CRISPR-associated endonuclease Cas1
		cas3	L06-129_01844	3,395	CRISPR-associated nuclease/helicase Cas3 subtype I-F/YPEST
		csy1	L06-129_01843	1,265	CRISPR-associated protein Csy1
		htpX_3	L06-129_01854	965	heat shock protein, protease HtpX
		0	L06-129_01853	470	Phosphate-starvation-inducible E
		0	L06-129_01848	569	small HspC2 heat shock protein

Overall, it becomes clear that our *L. pneumophila* isolates revealed high genome plasticity and adaptation to their environments and pathogenicity by acquisition and loss of genes and genomic islands taken together in the composition of all genomes and up to nine percent of the genome content comprising horizontally transferred genes (87).

3.17 Identification of pore-forming genes mediating cytotoxicity in *L. pneumophila* isolates

Central to the pore-forming mediated cytotoxicity of *L. pneumophila* are the Dot/Icm loci, which taken together directly assemble to a type IV secretion system (T4SS) (132, 133). Also, the toxin rtxA has an important role in the pore-mediated cytotoxicity (111, 134, 135). Although all *L. pneumophila* strains examined until today contain the complete Dot/Icm loci, sequence variations among the Dot/Icm genes among different *L. pneumophila* strains have been reported (132, 136).

Eleven Dot/Icm T4SS genes (*icmT* (72, 73), *icmS* (137), *icmR* (137), *icmQ* (137), *icmL/dotI* (112), *icmK/dotH* (112), *icmE/dotG* (112), *icmC/dotE* (112), *dotB* (112), *dotA* (112) and *icmW* (137)) and *rtxA* gene (135) are responsible for the pore-forming mediated cytotoxicity of *L. pneumophila*. Furthermore, three genes [legiolysin (*lly*) (138), alternative sigma factor RpoS (*ProA*) (136) and phospholipase (*plaB*) (139)] have hemolytic activity but do not display virulence effects on Red Blood Cells (RBCs) by *L. pneumophila*. Therefore, these three hemolytic but non-virulent genes were not studied. Dot/Icm T4SS pore-forming mediated cytotoxicity and rtxA genes were identified by a BLASTp search against the VFDB using *L. pneumophila* strain Philadelphia1 as default reference genome for the thirty-eight isolates which had been genome sequenced **(Table A6).**

Table A6 shows 75% to 100% similarity between the pore-forming mediated cytotoxicity genes. The *icmT* gene shared 82% similarity for all strains among twelve VFDB-annotated genes identified from *L. pneumophila* strains. The *IcmS* gene was shared by all strains (100%) except for cluster three (99%). For the *IcmR* gene the blastp similarity varied (from 91 to 97%) for all the strains. The *IcmQ* gene had a 100% similarity in clusters one, two and six plus a 99% similarity for clusters three, four and five. The *IcmL/DotI* genes shared 84% similarity for all strains on the average. The *IcmR/DotH, IcmE/DotG, IcmC/DotE, dotB, dotA* and *IcmW* gene shared (79-84%), (87-91%), (99-100%), (99-100%), (78-93%) and (98-100%) similarity for all strains, respectively. Also, the *rtxA* gene shared (75% to 85%) similarity for all *L. pneumophila* strains **(Figure 31).**

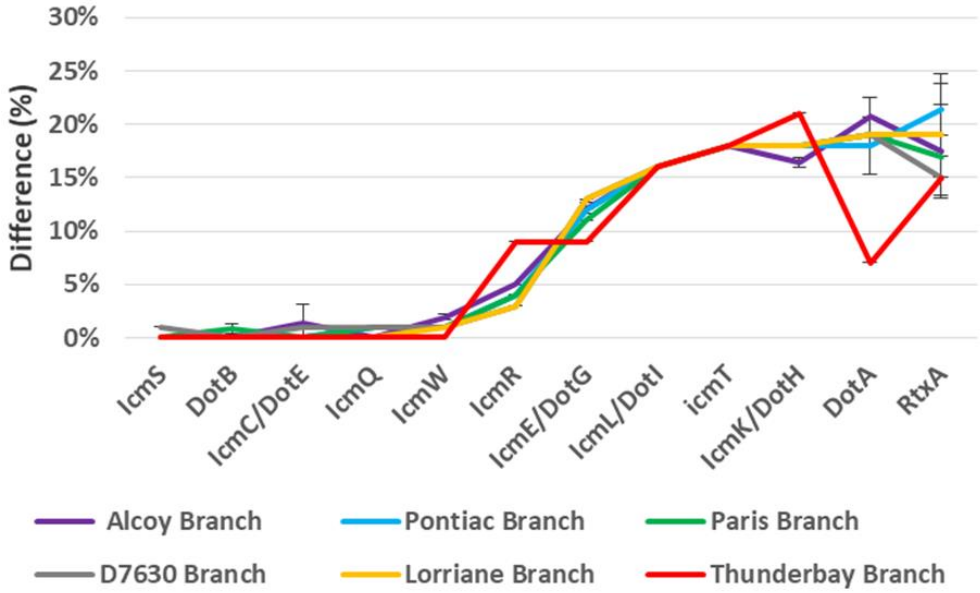


Figure 31: Percentages of the differences of the fifty-five *L. pneumophila* strains genome sequenced in the pore-forming mediated cytotoxicity virulence genes as identified by BLASTp search against the VFDB using *L. pneumophila* strain Philadelphia1 as default reference genome.

Taken together, this detailed analysis suggests that the twelve genes studied out of more than 200 of the Dot/Icm components of *L. pneumophila* are representing a large repertoire of effectors which are necessary for virulence (140). In general, all *L. pneumophila* strains shared the same Dot/Icm T4SS with less than 25% structural differences at the protein level.

4. Discussion

This is the first extensive study, which addresses the prevalence of *Legionella* spp in hospitals water systems in the West Bank. We set out to identify the levels of *Legionella* contamination in West Bank hospitals water supplies testing water samples and biofilms using both cultivation dependent and cultivation independent techniques targeting *Legionella* from genus level to clonal level for *L. pneumophila* using the MLVA-8(12) typing method.

4.1 *Legionella* abundance in water and biofilm

Culturable *L. pneumophila* were surprisingly low in water of the eight sampling sites in the West Bank. In only three of the eight sampling sites, *Legionella* were detected at all, and at only one site (F), *L. pneumophila* was detected more frequently in summer and autumn at a maximum value of 500 CFU/l. Hot water followed this trend and did not show higher concentrations than cold water. Using *L. pneumophila* specific PCR, *L. pneumophila* was detected at all sampling sites except site E. PCR detection was thus more sensitive with a broad range of site specific variability from 0 to 100% of water samples being positive on average per site. PCR-based detection was consistent with cultivation in a way that all cultivation positive samples were PCR positive. In biofilm, culturable *L. pneumophila* was detected at all sites, but with high sampling site specific variability. The incidence rate of culturable *L. pneumophila* in biofilm ranged from 3 to 30%. *L. pneumophila* specific PCR was positive on average for 40 to 93% of the biofilm samples per site.

Culture based detection was always lower compared to *L. pneumophila* specific PCR detection in bulk water, but culture based data were consistent with PCR based data, i.e. PCR based detection was always positive when cultivation was successful. This higher sensitivity of the PCR compared to cultivation was valid for water and biofilm. Culture based detection in biofilm was far more successful than in water. Culturable *L. pneumophila* was only detected in water when about 50% of the biofilm samples of the respective site and sampling were positive. The abundance of *L. pneumophila* by culture and PCR did not show a regional pattern. This was most obvious for the two sampling sites in Nablus (B, C) and Hebron (G, H), that differed by an order of magnitude. Compared to an annual study in a close climatic region, i.e. a water network at an university campus (Oranim) close to Haifa, much higher *Legionella* counts in drinking water were observed (49). The level of culturable *Legionella* in water at Oranim campus was in the range of 10 to 5800 CFU/l, with more than 60% of *Legionella* positive water samples. The prevalence of *L. pneumophila* in culture-based studies is in accordance with studies (50, 141) in Israel and Greece. Also, the prevalence of *L. pneumophila* in other studies was low (21.6% and 22%) in Kuwait and Tunisia respectively (52, 54). On the other hand, the prevalence of *L. pneumophila* was high (40%) in Jordan (142). Also in this study, samples with culturable *L. pneumophila* usually were obtained from the vicinity of culture positive biofilm samples.

A main cause of poor water quality is the build-up of biofilm in water distribution systems and tanks. *Legionella* are often considered as pioneers in creating biofilm in water distribution systems (100); with increasing biofilm formation the elimination of *Legionella* becomes difficult due to *L. pneumophila* tolerance to harsh physical and chemical conditions, including temperature and chlorination. Two buildings in hospital D water system; are relatively new wards established in 2010. However, *Legionella* was found in those two wards. A reasonable explanation is the problem in the control of water temperature which is maintained around 50°C, a suitable environment for the survival of *Legionella* within the hospital's water system. To prevent the accumulation of biofilm, a factor known to be associated with outbreaks of LD, water systems should be cleaned and maintained regularly including showerheads, shower hoses, faucets and main storage tanks.

The water reservoir of hospital C, E and H were analyzed for *Legionella spp*, and were found to be uncontaminated in most of water collections during the 2.3 years, whereas the swab from the distribution systems of the hospitals grew *Legionella*. Biofilms are found in every drinking water distribution system attached to the surface where they harbor many potentially pathogenic bacteria which are not isolated from the bulk water but are in the biofilms. The biofilm community is protected against adverse environmental conditions including disinfection (143-145).

4.2 Seasonal dynamics of *L. pneumophila*

Seasonal dynamics of *L. pneumophila* abundance showed a different pattern for water and biofilm. Abundance in water had a maximum in summer, and thus followed the temperature regime over the year. Abundance in biofilm increased from spring to summer and from summer to autumn indicating a maximum of *L. pneumophila* prevalence in biofilm in autumn. These seasonal patterns of water and biofilm were obtained by culture based analyses and by *L. pneumophila* specific PCR. In biofilm, the ratio of culturable *L. pneumophila* did approximately double from spring to autumn. From a seasonal perspective, there was no correlation between *L. pneumophila* abundance in water and biofilm due to their different seasonal maxima (146). In the water network at a campus close to Haifa, the highest fraction of *Legionella* positive samples (71%) were observed in spring and summer for both water and biofilm (49).

4.3 Influence of physico-chemical and bacteriological factors on *Legionella* abundance

In order to determine factors that influence the prevalence of *L. pneumophila* at the different sampling sites a broad set of physico-chemical and bacteriological parameters was analysed with respect to their relationship with *Legionella* abundance in water and biofilm of the eight sampling sites. In addition, the abundance of relevant *L. pneumophila* MLVA genotypes was included in this correlation analysis.

In terms of physico-chemical parameters only the magnesium concentration showed a significant negative correlation with *Legionella* abundance in water and biofilm that was observed by cultivation and PCR. Magnesium showed a very close negative correlation with the Ca/Mg ratio but was not correlated with the calcium concentration. Also, the calcium concentration did not correlate with *Legionella* abundance. Due to this Ca-Mg relationship, *Legionella* abundance showed a positive correlation with the Ca/Mg ratio. Therefore, we hypothesize that either Mg, the Ca/Mg ratio or a factor closely related to Mg had an influence on *Legionella* abundance in the drinking water of the West Bank.

In terms of bacteriological factors, there was no correlation of *Legionella* abundance in water and biofilm with heterotrophic plate counts, but with the prevalence of specific genotypes and their clonal complexes. The effect of heterotrophic plate counts (HPC) and physico-chemical parameters of water, on the presence and count of *Legionella* has been intensively studied in the past years. Nevertheless, some studies showed positive or contradicting association between *Legionella* counts with chlorine, temperature and HPC (49, 99, 147, 148). However, the magnesium concentration also showed a negative correlation with these specific genotypes and clonal complexes of *L. pneumophila* as it did on *Legionella* abundance.

The magnesium concentrations for all sampling sites were rather high ranging from 22 to 42 mg/l. High magnesium concentrations were observed in a hydrological study of spring and ground water in the West Bank with especially high values in the Eastern part of the West Bank (149). To the best of our knowledge, no study on *Legionella* abundance in drinking water is available that deals with drinking water with magnesium concentrations of this level.

4.4 *L. pneumophila* serogroups in the West Bank and comparison with worldwide view

L. pneumophila is the most pathogenic of *Legionella* spp, causing up to 90% of the cases of legionellosis (43, 44). Unfortunately, there are no previous data about legionellosis cases in the West Bank. According to the current epidemiological data available from the world, different *L. pneumophila* serogroups cause legionellosis. Overall, the great majority of strains isolated from the area under study in the West Bank were characterized as Sg.1 (62.3%). This fact followed the tendency already reported by other studies that have described Sg1 as the most frequently detected environmental isolate in different geographic regions (150-152). Besides the high prevalence of Sg1, other serogroups were isolated, where the fraction of non-Sg1 isolates went up to (37.7%). Sg6 was particularly abundant (16.7%). Sg8 and Sg10 were as well isolated, although in smaller proportions (3.3% and 0.7%, respectively). The results obtained here were highly concordant with two studies about the distribution of *L. pneumophila* serogroups in man-made water systems not related to human disease (152, 153). *L. pneumophila* Sg1 was the most frequently isolated serogroup followed by Sg6 in France and the UK, where they found as commonly

Sg10. Sg.6 was the serogroup responsible for cases of LD after Sg1, according to the European surveillance data (154) **(Table 3)**.

4.5 Biogeography of *L. pneumophila* genotype prevalence and their shaping force

MultiLocus Variable number of tandem Repeats Analysis (MLVA) was the method of choice in this study to evaluate the diversity and population structure of the large set of 180 *L. pneumophila* isolates in the West Bank due to its rapidity, cost, high reproducibility and discriminatory power (38, 39). Groundwater mostly is the origin of water supply in the West Bank (3). Interestingly, in the West Bank each hospital presented particular genotypes **(Figure 12 and Table 6).** Among the common genotypes, Gt4 (17) and Gt6 (18), the most frequent and more broadly distributed genotypes in each area belong to ST1 **(Table 5 and Figure A1).** Furthermore, ST1 is the most prevalent sequence type in the West Bank, and is the most dominant sequence type worldwide (152, 155-157). The high abundance of ST1 in the environment has been reported in several studies. In Japan, the majority of environmental isolates (29%) were ST1 (158) as well as in South Korea, where ST1 was distributed across all sampled facilities and regions and accounted for 48.1% of the isolates (159). ST1 was the most abundant sequence type among environmental isolates in Canada and it was found ubiquitously across the country (157). In a study conducted across the United States, ST1 was the most frequent sequence type among both clinical sporadic and environmental isolates, accounting for 25% and 49% of the total number of isolates respectively (156). In Europe, ST1 has also been reported as the most predominant sequence type among environmental isolates in Germany (88), England and Wales (152), Portugal(160) and Spain (161).

At the clonal level, the 180 *L. pneumophila* isolates were grouped according to their MLVA genotypes into four MLVA clonal complexes or VACCs (VACC1, VACC2, VACC5 and VACC11). VACC11 is defined for the first time by this study and it was formed exclusively by a group of 31 strains isolated from several hospitals. The population structure of *L. pneumophila* in the West Bank was characterized by the presence of a large cluster (VACC1, which included the “Paris” reference strain or ST1) that enclosed closely related Sg1 strains isolated from all of the sampling locations. VACC2, which included the “Philadelphia-1” reference strain. This pattern was consistent with previous studies (39). In the study by Visca *et al*., the population structure of a set of *L. pneumophila* strains with different origins from the European Union *Legionella* (EUL) collection was investigated by MLVA-8 genotyping. The majority of the strains (86%) were grouped into clonal complexes, of which VACC1 was the largest cluster followed by VACC2. The same pattern was observed among environmental strains isolated from water supply systems, cooling towers and hot springs in China (162). The majority of the isolates (89%) were enclosed in a large clonal complex formed by ST1 and single locus variants of this sequence type, which can be assigned to VACC1 **(Figure 12 and Figure A1)**.

A large set of 26 MLVA genotypes was retrieved from 180 *L. pneumophila* isolates with most isolates obtained from biofilm due to the low abundance of culturable *Legionella* strains in the bulk water. Only five strains were obtained from water: three from site F where the highest *Legionella* counts occurred in water, i.e. genotypes 10 (93) and 10 (141) belonging to VACC11; one from site A: Gt 16(1), VACC5; and one from site G: Gt 6(18), VACC1. The four genotypes obtained from water were represented by 5 to 30 isolates, and were frequently isolated from biofilm. The highest number of 74 strains was obtained for Gt 4(17) that was highly abundant in biofilm samples from the north of the West Bank, but was not obtained from water samples (97).

23 of the genotypes could be matched with 12 sequence types (Table 4) indicating that both genotyping methods revealed a high sub-species diversity of *L. pneumophila* isolates in the West Bank. The prevalence of the genotypes in the West Bank showed a site specific regional diversity. In the north of the West Bank a pronounced prevalence of Gt4(17) was observed. In the south, the genotype composition was more distinct from site to site and had endemic genotypes. Site G in the South was dominated by Gt6(18) that was endemic for this site, as was Gt10 (141) for site F.

As an index of genotype diversity per sampling site the ratio of the number of genotypes to the number of strains retrieved from the respective site was used. A correlation analysis showed that sites where high numbers of strains were isolated had a low genotype diversity compared to sites with low isolate numbers. Only from sites with low genotype diversity, isolates were obtained from water. This may indicate that higher numbers of *L. pneumophila* in the bulk water reduce the diversity in biofilm, or, that high *L. pneumophila* counts in water are often due to a rather restricted number of genotypes or even a single genotype.

4.6 Environmental drivers of genotype consortia or do *L. pneumophila* genotypes have specific niching preferences?

MLVA-genotypes with more than three representatives were analysed by average hierarchical cluster analysis and PCA with respect to their co-occurrence and positioning in the frame of bacteriological and physico-chemical parameters. Both statistical analyses showed the same three groupings of genotypes. Three to four genotypes formed one group. The environmental parameters and their statistical distinctness describing the niche of each group are given in **Table 8.**

The highest culturable *Legionella* abundance in water was associated with group A, comprising two VACC11-genotypes (Gt10(93), Gt10(141)) and one VACC2-genotype (Gt64(74)). This group was also characterized by low chloride and low magnesium concentrations compared to the other two groups. Group B comprised four genotypes, two of VACC1, the highly abundant Gt4(17) and Gt63(83), one genotype of VACC2 (Gt 13(72)) and one of VACC11 (Gt 9(93)). This group was distinct from group B and C with respect to most parameters, but was with most parameters in between the other two groups. Group B had the highest average Mg concentration at a high

variability. This may indicate that these genotypes are most tolerant to Mg. Group C comprised the VACC1-genotype Gt6(18), and the VACC5-genotypes Gt 16(1) and Gt40(47) that had a high prevalence in Hebron in the South of the West Bank. The environmental parameters of this group were characterized by high sulphate, chloride, calcium concentrations and a resulting high content of TDS.

4.7 Relevance of the observed genotypes for human health and environmental issues in the West Bank and worldwide

Only one previous study reported about the epidemiology of *L. pneumophila* in the West Bank hospitals water (57). In this current study, the systematic sampling of both water and biofilm carried out during 2.3 years in the West Bank lead to a large collection of 180 environmental *L. pneumophila* isolates that could bring new insights into the molecular epidemiology and genetic diversity of *L. pneumophila* in the West Bank. The studies, which focused on the genetic diversity of *L. pneumophila* in the environment, are not very common (161, 163) compared to those based on clinical isolates (153, 156, 164). The studies including environmental samples have been usually realized with the aim of establishing epidemiological links between clinical isolates and their presumed environmental sources, usually during legionellosis outbreaks (85, 93). However, since *Legionella* infections originate directly from the environment and do not occur by human-to-human transmission, the study of the genetic diversity and the distributions of *L. pneumophila* strains using this large set of environmental isolates could be relevant in public health control strategies in future epidemiological studies.

The genotypes isolated from the West Bank can be considered to be of environmental and human health relevance. As shown by Visca *et al*. (2011) and Sobral *et al*. (2011) (39, 85) MLVA genotype is well comparable to sequence based typing. In most cases, strains of a specific MLVA genotype have the same sequence type (ST). The applied MLVA using 13 loci can be considered to have a higher resolution than sequence based typing (85), meaning that several MLVA genotypes can be contained in one sequence type.

The most abundant genotype in the West Bank was Gt 4(17). This genotype comprises the pathogenic *L. pneumophila* strain Paris, belongs to sequence type 1 and serogroup 1 (40, 101). Gt 4(17) is distributed globally in anthropogenic freshwater systems (87). This strain is relevant as clinical isolate, and was occasionally related to outbreaks of LD (49, 162). More specifically, for the West Bank, some of the highly abundant genotypes can be considered to be of special health relevance. PCR-based direct analysis of the sequence types in respiratory specimens of pneumonia patients, revealed that genotypes belonging to ST1 and ST461 were present in half of the *L. pneumophila* positive specimens (165). We assume that the highly abundant representatives of these sequence types are the relevant pathogenic genotypes, means Gt4 (17) may be relevant for the detected ST1 infections, and the highly

abundant genotypes in the South of the West Bank , Gt 9 and 10 may be responsible for the ST461 infections (88).

The highly abundant genotypes Gt 4(17) and Gt 6(18) in the West Bank were also of high relevance in drinking water distribution systems and clinical isolates in Israel (18, 49, 86). Interestingly, none of the other genotypes in the West Bank were detected so far in Israel. In contrast to observations at the West Bank, genotypes Gt 4(17) and Gt 6(18) had a high abundance in the water and the biofilm in a campus drinking water distribution network in Northern Israel. The study comprised an annual cycle. The presence of Gt4 was associated with average *Legionella* counts in water of 2500 CFU/l at an average water temperature of 20.6°C. The presence of Gt6 was associated with average *Legionella* counts in water of 240 CFU/l at an average water temperature of 27.9°C. In the West Bank, Gt4(17) was only detected in biofilm in samples with no detection of culturable *Legionella* in bulk water; Gt6(18) was once isolated from water, at a single event of detectable *Legionella* in water in summer during a two annual cycles.

Rodriguez-Martinez *et al*. (2015) concluded that the presence of Gt4 could be considered as an indicator of high *Legionella* presence in drinking water, and suggested Gt4 as indicator genotype. Based on the observations in the West Bank where Gt4 was very frequently observed in biofilms in the Northern part without co-occurrence of high *Legionella* counts in water, we suggest that the presence in biofilm might not be an indicator for high *Legionella* counts in water (49). Due to the worldwide occurrence of Gt4 and the observations in Israel, we suggest that Gt4 may be regarded as an indicator of high *Legionella* abundance when showing up in the water phase. Furthermore, the presence of GT4 in biofilm might be considered as warning that, if conditions change, a "*Legionella* bloom" may be at risk.

4.8 In vitro-viulence traits of *L. pneumophila* genotypes

Little is known about the *L. pneumophila* in the West Bank in general and the virulence of these bacteria in particular. The understanding of *L. pneumophila* pathogenicity has been achieved mostly by the study of clinical isolates. Nevertheless, the analysis of the pathogenicity of environmental isolates is essential since environmental strains are considered the primary source of outbreaks as well as nosocomial and community acquired LD. In general, the degree of virulence of the of *L. pneumophila* strains is determined by studying various traits such as cytopathogenicity and infectivity within *A. castellanii* and THP-1 macrophages. Other approaches, as induction of pore-formation mediated cytotoxicity of the host and appearance of specific serogroup are also commonly used (110). Until now, none of these traits is decisive. Moreover, specific strains are thought to be particularly pathogenic since they are frequently found among clinical isolates, although it is still not clear which specific factors promote their pathogenicity. The assessment of the virulence of *L. pneumophila* strains is very supportive in order to understand the epidemiology of these bacteria. Therefore, to complement the genotyping results and better understand its

epidemiological significance, the virulence of representative MLVA genotypes and the virulence at clonal level were assessed for the genotypes. sixty environmental *L. pneumophila* isolates from the most abundant genotypes were tested for five different virulence assessment tests; infectivity of *A. castellanii*, cytopathogenicity of THP-1 macrophages, pore-forming mediated cytotoxicity, heat shock at 60°C for 30min and sodium chloride sensitivity.

Pore-forming mediated cytotoxicity and bacterial egress from the host cell upon termination of intracellular replication (69, 73) is one phase proposed of *Acanthamoeba* and macrophages killing by *L. pneumophila*. In this study, infectivity of *A. castellanii* and THP-1 macrophages and pore-forming mediated cytotoxicity of sRBC's were examined with intracellular replication in cells and presence of the Dot/Icm system to elucidate the virulence traits for different *L. pneumophila* genotypes and VACCs. Also, comparison between the WT AA100 and dotA mutant cultures were performed as control in three assays of virulence. The virulence of dotA mutant *L. pneumophila* was less than 15% for the three virulence tests. While, WT AA100 has high (60%) virulence activity for the tests. Moreover, *L. pneumophila* strain Paris and *L. pneumophila* strain Corby has higher virulence activity in comparison with *L. pneumophila* strain Philadelphia-1. Significant difference were shown for infection with *A. castellanii* (*P≤0.01*), THP-1 macrophages (*P≤0.05*) and pore forming mediated cytotoxicity of sRBC's (*P≤0.05*) **(Figure 19).**

The classification of the *L. pneumophila* strains used in this study into two main clusters provides important signs to the pathogenic features of *L. pneumophila* in the West Bank. Cluster one comprised the most abundant genotypes in the West Bank. The second cluster comprised the VACC in the West Bank. In this study, the infectivity assay either for *A. castellanii* or THP-1 macrophages had the same virulence distribution for the genotypes and clonal complexes. Gt6(18) which is also serogroup 1 and exclusively found in hospital G in Southern West Bank is the most virulent genotype. Also, VACC2 is the most virulent clonal complex **(Figure 20 and Figure 21)**. The virulence of *L. pneumophila* strains was tested by contact dependent pore formation mediated cytotoxicity of sRBC's. The Dot/Icm system is central for pathogenesis of *L. pneumophila*. Significant difference was clearly observed between *L. pneumophila* genotypes. *L. pneumophila* Gt4(17) showed the highest virulence activity (60%) which is Sg.1 corresponding to ST1. These results are in accordance with Kirby *et al* study (166) which indicates *L. pneumophila* Sg.1 as a highly virulent strain and ST1 is the most common Worldwide (Figure 7). Gt6(18) which is also Sg.1 and ST1 hemolysed sRBC's to (46%). Both genotypes [Gt4(17) and Gt6(18)] belong to VACC1. It is the most abundant clonal complex in the WB and worldwide and has the highest hemolytic activity (53%). *L. pneumophila* Gt12(84) shows high hemolytic activity (57%) which is Sg.8 corresponding to ST1358. For instance, ST1358 was among the most frequently isolated sequence types across Spain (167) and it is one of the virulent sequence types detected in a hotel outbreak in Calpe (Spain) affecting 44 patients including six deaths (94). Whereas, Gt9(92), Gt10(93) and Gt10(141) showed the low hemolytic activity (36%, 26% and 26% respectively) which corresponds to the VACC11.

VACC11 is the clonal complex identified for the first time in this study **(Figure 20 and Figure 21)** and has the least clonal complex hemolytic activity (29%). The three least hemolytic activity genotypes [Gt10(93), Gt10(141) and Gt64(74)] are site related in hospital F **(Figure 20 and Figure 21)**. Also, *L. pneumophila* str. Paris and *L. pneumophila* str. Corby have the highest virulence activity (75% and 61% respectively). They also belong to VACC1. VACC2 has hemolytic activity of (41%) and the hemolytic activity of the *L. pneumophila* str. Pheladelphia-1 was (34%) belongs also to VACC2 **(Figure 21)**.

A pore forming mediated cytotoxicity of *L. pneumophila* is associated with *L. pneumophila* virulence. As when MOI increase, *L. pneumophila* insert a pore which leads to osmotic lysis and replication death (112). Normally, *L. pneumophila* insert pores into its phagosomal membrane to block fusion with the lysosomal network (166). Instead, the cytotoxin could facilitate the lysis of the host membranes by *L. pneumophila* (67). Pore forming cytotoxicity plays an important role in the cytopathogenicity and cytotoxicity of *L. pneumophila* isolates since all the isolates have virulent activity. In fact, this is not only factor responsible for the virulence in *L. pneumophila*. Feeley *et al* observed that *Legionella* were sensitive to sodium ions. So, potassium hydroxide is recommended to adjust the pH of BCYE (168). The genetic studies showed that Dot/Icm loci is the most important loci responsible for sodium tolerance (169). Also, ompS gene is negatively regulated by inhibitory level of sodium chloride (170). The heating of water to temperatures of 60°C or more is widely used to reduce total bacterial numbers and inactivate the number of pathogens like *Legionella* (17, 171).

As shown in **(Figure 20)** no significant difference was present between *L. pneumophila* genotypes. Gt4(17) tolerated sodium chloride (12%) more than the other genotypes. However, Gt64(74) was the most resistant to 60°C heat shock test. Both Gt(4(17) and Gt6(18) belong to VACC1. As shown in **(Figure 21)** significant difference ($P \leq 0.05$) was shown between VACC1 and VACC5 in sodium tolerance test. Also VACC1 is significantly resistant to 60°C heat than VACC5 ($P \leq 0.05$). As shown in **(Figure 22)** environmental niches don't have clear effect on *L. pneumophila* virulence activity. In general *L. pneumophila* responds to amino acid depletion by switching from a replicative phase to a virulent form, a state characterized by cytotoxicity assays and sodium sensitivity (67).

4.9 Comparison of taxonomic resolution of MLVA genotyping and SNP analysis

In the present study, we sequenced and compared the genomes of thirty-eight *L. pneumophila* isolates covering different serogroups, MLVA-8(12) genotypes, isolation sources (clinical and environmental) from the West Bank and Germany. The majority of previous studies have used WGS to study genomics of *L. pneumophila* and to discriminate outbreak strains of *L. pneumophila* retrospectively or during the outbreak (89, 172). Moran-Gilad el al (172) studied outbreak isolates using core genome multi-

locus sequence typing (cgMLST) and observed that cgMLST has a high enough resolution to identify *L. pneumophila* outbreak strains (172). Another study compared whole genome SNP analysis with sequence based typing (SBT) for genotyping *L. pneumophila* and they found a comparable relationship between genome SNPs and SBT (89). Qin *et al* defined the minimum core genome (MCG) and demonstrated that MCG typing enabled differentiating *L. pneumophila* strain into groups having major differences in virulence and phenotypic features (95).

Actually, a threshold for the number of SNPs necessary to identify outbreak-associated *L. pneumophila* strains still has to be established. For example, more than 200 SNPs have been reported in phylogenetically closely related *L. pneumophila* outbreak strains (93, 173). Mercante *et al* (174) showed that up to 20 core SNPs were identified in comparison of Philadelphia clade *L. pneumophila* isolates. In the present study, up to 120 SNPs were identified in the same MLVA-8(12) genotype from the same sub-lineage in a phylogenomic cluster. In particular, the environmental isolates A156_Gt64(74)_Ps and A129_Gt64(74)_Ps Gt64(74) from a biofilm of hospital F in the Bethlehem area were identical for all reference genomes except for *L. pneumophila* strain Lens and *L. pneumophila* strain LPE509 differed by only one and four SNPs, respectively. The environmental isolates H29_Gt22(100)_D, H34_Gt22(102)_D and H35_Gt22(102)_D from the HZI differed by less than 70 SNPs. A sub-lineage of cluster six for the clinical isolates L06-153_Cl_Gt71(135)_D and L06-129_Cl_Gt71(135)_D from Brandenburg, Germany differed by only three SNPs against all reference genomes used in this study. On the other hand, the cluster four sub-lineage of Gt4(17) contained four strains (two reference genomes from the strains Paris and OLDA from France and USA, respectively and two environmental isolates A5_Gt4(17)_Ps and A139_Gt4(17)_Ps from Jerusalem and Nablus, West Bank) differed by less than 260 SNPs against all reference genomes in this study. These results are in concordance with Khodr *et al* (101). They sequenced six ST1 genomes (four clinical and environmental isolates from a hospital and the other two were unrelated) and observed that geographically unrelated isolates differed by more than 1,500 SNPs. In comparison our hospital isolates differed by up to 20 SNPs between two strains. Several previous studies showed that *L. pneumophila* strains have the same genotype (either SBT or MLVA) isolated from the same area have less SNP differences than strains isolated from different areas. This demonstrates that the geographical origin is as important as the genotype (**Figure 23** and **Table A4**).

Morozova *et al* (175) showed that the Dot/Icm genes are highly conserved in *L. pneumophila* strains. After whole genome sequencing (WGS) technology was available on the market, Gomez-Valero *et al* (140, 176) confirmed the previous study showing high conservation (98%) among orthologs of the reference strains Corby, Paris, Philadelphia and Lens with few exceptions in the *dotA* gene. The *dotA* gene is an essential gene for virulence activity of *L. pneumophila* strains since it encodes an integral membrane protein with eight domains. This explains why a *dotA* mutant of *L. pneumophila* strain Corby is being used as a negative control for all virulence assays (177). Costa *et al* (178) analyzed 300 *dotA* gene sequences from *L. pneumophila*

strains and demonstrated that pathogenic *L. pneumophila* strains belong to a subset of the genotypes existing in the environment. Khodr *et al* (101) explained the high variation of the *dotA* gene of *L. pneumophila* by indicating that this gene is a target for host speciation and adaptive evolution to different hosts and environments. Dumenil *et al* (179) showed that *IcmR* is a regulator gene for the *IcmQ* gene that possesses pore-forming activity. In addition, Gomez-Valero *et al* (180) demonstrated that *dotB, IcmS* and *IcmW* are highly conserved genes. These facts are in accordance with our results (**Table A6 and Figure 31**) showing that *IcmR* (91-96%), *IcmS* and *dotB* (99-100%) and *IcmQ* (98-100%) were highly conserved gene while *dotA* had only a 78% to 93% gene similarity for our *L. pneumophila* strains. Overall, the Dot/Icm system is a highly conserved and complex molecular system.

4.10 Evolutionary scenarios explaining levels of diversity observed in virulence and other genes

L. pneumophila shows large genome plasticity and possesses a highly dynamic accessory genome due to its peculiar evolutionary conditions as an internal parasite, its substantial horizontal gene transfer (HGT) and the presence of mobile genetic elements (92). Mobile genetic elements together with bacteriophages and plasmids are called integrative conjugative elements (181). Most of *L. pneumophila* mobile genetic elements are integrative conjugative elements encoding different T4SS which mainly consist of the *L. pneumophila* vir homologues regions (Lvh-region), the Trb/Tra family of conjugative elements and other genomic islands associated with the T4SS gene family. The Lvh-region consists of eleven genes encoding a T4SS and are located on DNA islands with high CG content (140, 182). The Lvh T4ASS is not present in *L. pneumophila* strain Corby, but a similar T4ASS is integrated in this site (tmRNA) and a second genomic island carrying a T4ASS is integrated in the tRNA gene, a site not occupied by a mobile genetic element in the strains Paris, Lens or Philadelphia (182).

In addition to HGT, gene duplication contributes to the evolution of bacterial genomes (101). It is one of the main driving forces of genetic diversity, adaptation to specific environments and speciation. Four paralogues genes (sdeA, sdeB, sdeC and sdeE) are on example of gene duplication in the expanded Dot/Icm effector repertoire (40, 183). Khodr el al (101) summarized the distribution of integrative conjugative element and essential duplication genes present in *L. pneumophila* strain. These genomic elements could be observed in our study in the isolates A129_Gt64(74)_Ps, A156_Gt64(74)_Ps, H29_Gt22(100)_D, H34_Gt22(102)_D, H35_Gt22(102)_D, A193_Gt40(47)_Ps, A195_Gt40(47)_Ps and L04-545_Cl_Gt40(47)_D (**Appendices A7-A9**). In this study, most of the T4SS mobile genetic elements or duplication genes do not express a strong virulence phenotype.

According to David *et al* (87) a certain local microevolution could be observed if isolates from the same site have been obtained at different times. Such a case of micro-evolution could have occurred in the set of isolates obtained from the HZI hot water supply. Strain H23_Gt14(30)_D of Gt14(30) has had more time to diversify by

genetic drift since it was isolated in 2013 in contrast to the isolates H1_Gt14(31)_D, H2_Gt14(31)_D, and H3_Gt14(31)_D obtained in 2009. It is also possible that *L. pneumophila* strain was subjected to different selection pressures during its life in the hot water supply system. If this is the case then a likely explanation is the harsh conditions in the hot water selected for changes in the metabolism of the isolates making them able to adapt to this environment (184). This selection process could make *L. pneumophila* more virulent because it adapted better to high temperature occurring in the human body. Differentiating between these putative evolutionary scenarios will be difficult and will require a greater understanding of the effect of diversity within hot water samples and isolates from this environment (**Table 10**). To this end, further genomic and metagenomics studies are needed using these materials collected during this time.

5. References

1. The World Factbook- Middle East: West Bank. Central Intelligence Agency 2016.
2. Koek E. Water for one people only- Discriminatory access and 'Water-Apartheid' in the OPT. Al-Haq. 2013.
3. Authority PW. STATUS REPORT OF WATER RESOURCES IN THE OCCUPIED STATE OF PALESTINE-2012. Annual Water Resources Status Report 2012. October,2013:1-17.
4. Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, et al. Legionnaires' disease: description of an epidemic of pneumonia. N Engl J Med. 1977;297(22):1189-97.
5. McDade JE, Shepard CC, Fraser DW, Tsai TR, Redus MA, Dowdle WR. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N Engl J Med. 1977;297(22):1197-203.
6. Muder RR, Yu VL. Infection due to Legionella species other than *L. pneumophila*. Clin Infect Dis. 2002;35(8):990-8.
7. Pearce MM, Theodoropoulos N, Mandel MJ, Brown E, Reed KD, Cianciotto NP. *Legionella cardiaca* sp. nov., isolated from a case of native valve endocarditis in a human heart. Int J Syst Evol Microbiol. 2012;62(Pt 12):2946-54.
8. Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. Ecological distribution of *Legionella pneumophila*. Appl Environ Microbiol. 1981;41(1):9-16.
9. Murga R, Forster TS, Brown E, Pruckler JM, Fields BS, Donlan RM. Role of biofilms in the survival of *Legionella pneumophila* in a model potable-water system. Microbiology. 2001;147(Pt 11):3121-6.
10. Keevil CW. Rapid detection of biofilms and adherent pathogens using scanning confocal laser microscopy and episcopic differential interference contrast microscopy. Water Sci Technol. 2003;47(5):105-16.
11. Declerck P, Behets J, van Hoef V, Ollevier F. Detection of *Legionella* spp. and some of their amoeba hosts in floating biofilms from anthropogenic and natural aquatic environments. Water Res. 2007;41(14):3159-67.
12. Rowbotham TJ. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. J Clin Pathol. 1980;33(12):1179-83.
13. Taylor M, Ross K, Bentham R. *Legionella*, protozoa, and biofilms: interactions within complex microbial systems. Microbial ecology. 2009;58(3):538-47.
14. Barker J, Brown MR, Collier PJ, Farrell I, Gilbert P. Relationship between *Legionella pneumophila* and *Acanthamoeba polyphaga*: physiological status and susceptibility to chemical inactivation. Appl Environ Microbiol. 1992;58(8):2420-5.
15. Cirillo JD, Cirillo SL, Yan L, Bermudez LE, Falkow S, Tompkins LS. Intracellular growth in *Acanthamoeba castellanii* affects monocyte entry mechanisms and enhances virulence of Legionella pneumophila. Infect Immun. 1999;67(9):4427-34.
16. Bigot R, Bertaux J, Frere J, Berjeaud JM. Intra-amoeba multiplication induces chemotaxis and biofilm colonization and formation for *Legionella*. PloS one. 2013;8(10):e77875.
17. Wadowsky RM, Wolford R, McNamara AM, Yee RB. Effect of temperature, pH, and oxygen level on the multiplication of naturally occurring *Legionella pneumophila* in potable water. Appl Environ Microbiol. 1985;49(5):1197-205.
18. Sharaby Y, Rodriguez-Martinez S, Oks O, Pecellin M, Mizrahi H, Peretz A, et al. Temperature-Dependent Growth Modeling of Environmental and Clinical *Legionella*

pneumophila Multilocus Variable-Number Tandem-Repeat Analysis (MLVA) Genotypes. Appl Environ Microbiol. 2017;83(8).
19. Stout J, Yu VL, Vickers RM, Zuravleff J, Best M, Brown A, et al. Ubiquitousness of *Legionella pneumophila* in the water supply of a hospital with endemic Legionnaires' disease. N Engl J Med. 1982;306(8):466-8.
20. Borella P, Montagna MT, Stampi S, Stancanelli G, Romano-Spica V, Triassi M, et al. *Legionella* contamination in hot water of Italian hotels. Appl Environ Microbiol. 2005;71(10):5805-13.
21. Fields BS, Benson RF, Besser RE. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev. 2002;15(3):506-26.
22. Swanson MS, Hammer BK. *Legionella pneumophila* pathogesesis: a fateful journey from amoebae to macrophages. Annu Rev Microbiol. 2000;54:567-613.
23. Mauchline WS, James BW, Fitzgeorge RB, Dennis PJ, Keevil CW. Growth temperature reversibly modulates the virulence of *Legionella pneumophila*. Infect Immun. 1994;62(7):2995-7.
24. Fiore AE, Nuorti JP, Levine OS, Marx A, Weltman AC, Yeager S, et al. Epidemic Legionnaires' disease two decades later: old sources, new diagnostic methods. Clin Infect Dis. 1998;26(2):426-33.
25. Pedro-Botet ML, Stout JE, Yu VL. Legionnaires' disease contracted from patient homes: the coming of the third plague? Eur J Clin Microbiol Infect Dis. 2002;21(10):699-705.
26. Breiman RF, Fields BS, Sanden GN, Volmer L, Meier A, Spika JS. Association of shower use with Legionnaires' disease. Possible role of amoebae. JAMA. 1990;263(21):2924-6.
27. Shin S. Innate Immunity to Intracellular Pathogens: Lessons Learned from *Legionella pneumophila*. Advances in applied microbiology. 2012;79:43-71.
28. Correia AM, Ferreira JS, Borges V, Nunes A, Gomes B, Capucho R, et al. Probable Person-to-Person Transmission of Legionnaires' Disease. N Engl J Med. 2016;374(5):497-8.
29. Lo Presti F, Riffard S, Meugnier H, Reyrolle M, Lasne Y, Grimont PA, et al. *Legionella gresilensis* sp. nov. and *Legionella beliardensis* sp. nov., isolated from water in France. Int J Syst Evol Microbiol. 2001;51(Pt 6):1949-57.
30. Riffard S, Lo Presti F, Normand P, Forey F, Reyrolle M, Etienne J, et al. Species identification of *Legionella* via intergenic 16S-23S ribosomal spacer PCR analysis. International journal of systematic bacteriology. 1998;48 Pt 3:723-30.
31. Hookey JV, Birtles RJ, Saunders NA. Intergenic 16S rRNA gene (rDNA)-23S rDNA sequence length polymorphisms in members of the family Legionellaceae. J Clin Microbiol. 1995;33(9):2377-81.
32. Fry NK, Afshar B, Bellamy W, Underwood AP, Ratcliff RM, Harrison TG. Identification of *Legionella* spp. by 19 European reference laboratories: results of the European Working Group for *Legionella* Infections External Quality Assessment Scheme using DNA sequencing of the macrophage infectivity potentiator gene and dedicated online tools. Clin Microbiol Infect. 2007;13(11):1119-24.
33. Ratcliff RM, Lanser JA, Manning PA, Heuzenroeder MW. Sequence-based classification scheme for the genus *Legionella* targeting the mip gene. J Clin Microbiol. 1998;36(6):1560-7.
34. Stolhaug A, Bergh K. Identification and differentiation of *Legionella pneumophila* and *Legionella* spp. with real-time PCR targeting the 16S rRNA gene and species identification by mip sequencing. Appl Environ Microbiol. 2006;72(9):6394-8.
35. Ko KS, Lee HK, Park MY, Lee KH, Yun YJ, Woo SY, et al. Application of RNA polymerase beta-subunit gene (rpoB) sequences for the molecular differentiation of *Legionella* species. J Clin Microbiol. 2002;40(7):2653-8.

36. Grattard F, Ginevra C, Riffard S, Ros A, Jarraud S, Etienne J, et al. Analysis of the genetic diversity of *Legionella* by sequencing the 23S-5S ribosomal intergenic spacer region: from phylogeny to direct identification of isolates at the species level from clinical specimens. Microbes and infection. 2006;8(1):73-83.
37. Fry NK, Alexiou-Daniel S, Bangsborg JM, Bernander S, Castellani Pastoris M, Etienne J, et al. A multicenter evaluation of genotypic methods for the epidemiologic typing of *Legionella pneumophila* serogroup 1: results of a pan-European study. Clin Microbiol Infect. 1999;5(8):462-77.
38. Pourcel C, Visca P, Afshar B, D'Arezzo S, Vergnaud G, Fry NK. Identification of variable-number tandem-repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme. J Clin Microbiol. 2007;45(4):1190-9.
39. Visca P, D'Arezzo S, Ramisse F, Gelfand Y, Benson G, Vergnaud G, et al. Investigation of the population structure of *Legionella pneumophila* by analysis of tandem repeat copy number and internal sequence variation. Microbiology. 2011;157(Pt 9):2582-94.
40. Cazalet C, Rusniok C, Bruggemann H, Zidane N, Magnier A, Ma L, et al. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. Nat Genet. 2004;36(11):1165-73.
41. Glick TH, Gregg MB, Berman B, Mallison G, Rhodes WW, Jr., Kassanoff I. Pontiac fever. An epidemic of unknown etiology in a health department: I. Clinical and epidemiologic aspects. Am J Epidemiol. 1978;107(2):149-60.
42. Marston BJ, Lipman HB, Breiman RF. Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality. Arch Intern Med. 1994;154(21):2417-22.
43. Benin AL, Benson RF, Besser RE. Trends in legionnaires disease, 1980-1998: declining mortality and new patterns of diagnosis. Clin Infect Dis. 2002;35(9):1039-46.
44. Yu VL, Plouffe JF, Pastoris MC, Stout JE, Schousboe M, Widmer A, et al. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. J Infect Dis. 2002;186(1):127-8.
45. (ECDC) ECfLDPaCS. ECDC Surveillance Report: Legionnaires' Disease in Europe. ECDC Surveillance Report. 2013.
46. Ben-Dror G, Mizerizky Y, Viar G, Zuker M, Miron D. [The epidemiology and clinical features of *Legionella pneumonia* (LP) in patients older than 60 years old who were hospitalized with pneumonia in northern Israel]. Harefuah. 2002;141(8):680-2, 763.
47. Oren I, Zuckerman T, Avivi I, Finkelstein R, Yigla M, Rowe JM. Nosocomial outbreak of *Legionella pneumophila* serogroup 3 pneumonia in a new bone marrow transplant unit: evaluation, treatment and control. Bone Marrow Transplant. 2002;30(3):175-9.
48. Moran-Gilad J, Mentasti M, Lazarovitch T, Huberman Z, Stocki T, Sadik C, et al. Molecular epidemiology of Legionnaires' disease in Israel. Clin Microbiol Infect. 2014;20(7):690-6.
49. Rodriguez-Martinez S, Sharaby Y, Pecellin M, Brettar I, Höfle M, Halpern M. Spatial distribution of *Legionella pneumophila* MLVA-genotypes in a drinking water system. Water Res. 2015;77:119-32.
50. Rachel Yarom RSaRA. *Legionella pneumophila* serogroup 3 prevalence in drinking water survey in Israel (2003 –2007). Water Science & Technology: Water Supply—WSTWS. 2010;10.5:746-52.
51. Erdogan H, Arslan H. Colonization of *Legionella* species in Turkish baths in hotels in Alanya, Turkey. Environmental monitoring and assessment. 2015;187(5):235.
52. Ghrairi T, Chaftar N, Jarraud S, Berjeaud JM, Hani K, Frere J. Diversity of *legionellae* strains from Tunisian hot spring water. Research in microbiology. 2013;164(4):342-50.

53. Mekkour M, Ben Driss E, Tai J, Squinazi F, Forey F, Jarraud S, et al. molecular typing of *Legionella pneumophila* strains isolated from environment in Morocco. Cellular and molecular biology. 2012;58 Suppl:OL1709-14.
54. Al-Matawah QA, Al-Zenki SF, Qasem JA, Al-Waalan TE, Ben Heji AH. Detection and Quantification of *Legionella pneumophila* from Water Systems in Kuwait Residential Facilities. Journal of pathogens. 2012;2012:138389.
55. Al-Matawah Q, Al-Zenki S, Al-Azmi A, Al-Waalan T, Al-Salameen F, Hejji AB. *Legionella* detection and subgrouping in water air-conditioning cooling tower systems in Kuwait. Environmental science and pollution research international. 2015;22(13):10235-41.
56. (ECDC) ECfLDPaCS. Rapid risk assessment. Increase of cases of Legionnaires' disease in EU travellers returning from Dubai, October-December 2016. ECDC Report. 2016.
57. Shareef A, Mimi Z. The Hospital Tap Water System as a Source of Nosocomial Infections for Staff Members and Patients in West Bank Hospitals. Environmental Forensics. 2008;9(6):226-30
58. Marrie TJ, Haldane D, MacDonald S, Clarke K, Fanning C, Le Fort-Jost S, et al. Control of endemic nosocomial legionnaires' disease by using sterile potable water for high risk patients. Epidemiol Infect. 1991;107(3):591-605.
59. Sabria M, Pedro-Botet ML, Gomez J, Roig J, Vilaseca B, Sopena N, et al. Fluoroquinolones vs macrolides in the treatment of Legionnaires disease. Chest. 2005;128(3):1401-5.
60. Lück C, Fry NK, Helbig JH, Jarraud S, Harrison TG. Typing methods for *legionella*. Methods in molecular biology. 2013;954:119-48.
61. Steinert M, Hentschel U, Hacker J. *Legionella pneumophila*: an aquatic microbe goes astray. FEMS Microbiol Rev. 2002;26(2):149-62.
62. Mercante JW, Winchell JM. Current and emerging *Legionella* diagnostics for laboratory and outbreak investigations. Clin Microbiol Rev. 2015;28(1):95-133.
63. Kirschner AK, Rameder A, Schrammel B, Indra A, Farnleitner AH, Sommer R. Development of a new CARD-FISH protocol for quantification of *Legionella pneumophila* and its application in two hospital cooling towers. Journal of applied microbiology. 2012;112(6):1244-56.
64. Mizrahi H, Peretz A, Lesnik R, Aizenberg-Gershtein Y, Rodriguez-Martinez S, Sharaby Y, et al. Comparison of sputum microbiome of legionellosis-associated patients and other pneumonia patients: indications for polybacterial infections. Scientific reports. 2017;7:40114.
65. Delafont V, Brouke A, Bouchon D, Moulin L, Hechard Y. Microbiome of free-living amoebae isolated from drinking water. Water Res. 2013;47(19):6958-65.
66. Fields BS. The molecular ecology of *legionellae*. Trends Microbiol. 1996;4(7):286-90.
67. Byrne B, Swanson MS. Expression of *Legionella pneumophila* virulence traits in response to growth conditions. Infect Immun. 1998;66(7):3029-34.
68. Gao LY, Harb OS, Abu Kwaik Y. Utilization of similar mechanisms by *Legionella pneumophila* to parasitize two evolutionarily distant host cells, mammalian macrophages and protozoa. Infect Immun. 1997;65(11):4738-46.
69. Alli OA, Gao LY, Pedersen LL, Zink S, Radulic M, Doric M, et al. Temporal pore formation-mediated egress from macrophages and alveolar epithelial cells by *Legionella pneumophila*. Infect Immun. 2000;68(11):6431-40.
70. Horwitz MA, Silverstein SC. Legionnaires' disease bacterium (*Legionella pneumophila*) multiples intracellularly in human monocytes. J Clin Invest. 1980;66(3):441-50.
71. Hilbi H, Segal G, Shuman HA. Icm/dot-dependent upregulation of phagocytosis by *Legionella pneumophila*. Mol Microbiol. 2001;42(3):603-17.

72. Bitar DM, Molmeret M, Kwaik YA. Structure-function analysis of the C-terminus of IcmT of *Legionella pneumophila* in pore formation-mediated egress from macrophages. FEMS Microbiol Lett. 2005;242(1):177-84.
73. Molmeret M, Alli OA, Zink S, Flieger A, Cianciotto NP, Kwaik YA. icmT is essential for pore formation-mediated egress of *Legionella pneumophila* from mammalian and protozoan cells. Infect Immun. 2002;70(1):69-78.
74. Li XH, Zeng YL, Gao Y, Zheng XC, Zhang QF, Zhou SN, et al. The ClpP protease homologue is required for the transmission traits and cell division of the pathogen *Legionella pneumophila*. BMC Microbiol. 2010;10:54.
75. Hammer BK, Swanson MS. Co-ordination of *legionella pneumophila* virulence with entry into stationary phase by ppGpp. Mol Microbiol. 1999;33(4):721-31.
76. Yu VL, Stout JE. Hospital characteristics associated with colonization of water systems by *Legionella* and risk of nosocomial legionnaires' disease: a cohort study of 15 hospitals. Infect Control Hosp Epidemiol. 2000;21(7):434-5.
77. Schijven JF, de Roda Husman AM. Effect of climate changes on waterborne disease in The Netherlands. Water Sci Technol. 2005;51(5):79-87.
78. Organization IS. Water quality-detection and enumeration of *Legionella*. International standard ISO, 11731, International Standards Organization. 1998;Geneva, Switzerland.
79. Steinert M, Emody L, Amann R, Hacker J. Resuscitation of viable but nonculturable *Legionella pneumophila* Philadelphia JR32 by Acanthamoeba castellanii. Appl Environ Microbiol. 1997;63(5):2047-53.
80. Nederbragt AJ, Balasingham A, Sirevag R, Utkilen H, Jakobsen KS, Anderson-Glenna MJ. Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis. J Microbiol Methods. 2008;73(2):111-7.
81. Villari P, Motti E, Farullo C, Torre I. Comparison of conventional culture and PCR methods for the detection of *Legionella pneumophila* in water. Letters in applied microbiology. 1998;27(2):106-10.
82. Bowden KE, Williams MM, Cassiday PK, Milton A, Pawloski L, Harrison M, et al. Molecular epidemiology of the pertussis epidemic in Washington State in 2012. J Clin Microbiol. 2014;52(10):3549-57.
83. Kjeldsen MK, Torpdahl M, Campos J, Pedersen K, Nielsen EM. Multiple-locus variable-number tandem repeat analysis of *Salmonella enterica* subsp. *enterica* serovar Dublin. Journal of applied microbiology. 2014;116(4):1044-54.
84. Roussel S, Felix B, Vingadassalon N, Grout J, Hennekinne JA, Guillier L, et al. *Staphylococcus aureus* strains associated with food poisoning outbreaks in France: comparison of different molecular typing methods, including MLVA. Frontiers in microbiology. 2015;6:882.
85. Sobral D, Le Cann P, Gerard A, Jarraud S, Lebeau B, Loisy-Hamon F, et al. High-throughput typing method to identify a non-outbreak-involved *Legionella pneumophila* strain colonizing the entire water supply system in the town of Rennes, France. Appl Environ Microbiol. 2011;77(19):6899-907.
86. Sharaby Y, Rodriguez-Martinez S, Pecellin M, Sela R, Peretz A, Höfle M, et al. Virulence traits of environmental and clinical *Legionella pneumophila* MLVA genotypes. Appl Environ Microbiol. 2018.
87. David S, Rusniok C, Mentasti M, Gomez-Valero L, Harris SR, Lechat P, et al. Multiple major disease-associated clones of *Legionella pneumophila* have emerged recently and independently. Genome Res. 2016;26(11):1555-64.

88. Borchardt J, Helbig JH, Lück PC. Occurrence and distribution of sequence types among *Legionella pneumophila* strains isolated from patients in Germany: common features and differences to other regions of the world. Eur J Clin Microbiol Infect Dis. 2008;27(1):29-36.
89. Underwood AP, Jones G, Mentasti M, Fry NK, Harrison TG. Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. BMC Microbiol. 2013;13:302.
90. Bosch T, Euser SM, Landman F, Bruin JP, EP IJ, den Boer JW, et al. Whole-Genome Mapping as a Novel High-Resolution Typing Tool for *Legionella pneumophila*. J Clin Microbiol. 2015;53(10):3234-8.
91. Chien M, Morozova I, Shi S, Sheng H, Chen J, Gomez SM, et al. The genomic sequence of the accidental pathogen *Legionella pneumophila*. Science. 2004;305(5692):1966-8.
92. Gomez-Valero L, Buchrieser C. Genome dynamics in *Legionella*: the basis of versatility and adaptation to intracellular replication. Cold Spring Harbor perspectives in medicine. 2013;3(6).
93. Reuter S, Harrison TG, Koser CU, Ellington MJ, Smith GP, Parkhill J, et al. A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. BMJ open. 2013;3(1).
94. Sanchez-Buso L, Guiral S, Crespi S, Moya V, Camaro ML, Olmos MP, et al. Genomic Investigation of a Legionellosis Outbreak in a Persistently Colonized Hotel. Frontiers in microbiology. 2015;6:1556.
95. Qin T, Zhang W, Liu W, Zhou H, Ren H, Shao Z, et al. Population structure and minimum core genome typing of *Legionella pneumophila*. Scientific reports. 2016;6:21356.
96. Gaia V, Fry NK, Harrison TG, Peduzzi R. Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. J Clin Microbiol. 2003;41(7):2932-9.
97. Pecellín M. Structure and virulence of *Legionella pneumophila* populations from freshwater systems in Germany and Middle East. Technische Universität Braunschweig. 2016;Braunschweig, Germany.
98. Bargellini A, Marchesi I, Righi E, Ferrari A, Cencetti S, Borella P, et al. Parameters predictive of *Legionella* contamination in hot water systems: association with trace elements and heterotrophic plate counts. Water Res. 2011;45(6):2315-21.
99. Rakic A, Peric J, Foglar L. Influence of temperature, chlorine residual and heavy metals on the presence of *Legionella pneumophila* in hot water distribution systems. Annals of agricultural and environmental medicine : AAEM. 2012;19(3):431-6.
100. Schwake DO, Alum A, Abbaszadegan M. Impact of environmental factors on *legionella* populations in drinking water. Pathogens. 2015;4(2):269-82.
101. Khodr A, Kay E, Gomez-Valero L, Ginevra C, Doublet P, Buchrieser C, et al. Molecular epidemiology, phylogeny and evolution of *Legionella*. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2016;43:108-22.
102. Steinert M, Heuner K, Buchrieser C, Albert-Weissenberger C, Glöckner G. *Legionella* pathogenicity: genome structure, regulatory networks and the host cell response. Int J Med Microbiol. 2007;297(7-8):577-87.
103. D'Auria G, Jimenez-Hernandez N, Peris-Bondia F, Moya A, Latorre A. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. BMC genomics. 2010;11:181.
104. Khan MA, Knox N, Prashar A, Alexander D, Abdel-Nour M, Duncan C, et al. Comparative Genomics Reveal That Host-Innate Immune Responses Influence the Clinical Prevalence of *Legionella pneumophila* Serogroups. PloS one. 2013;8(6):e67298.

105. Mercante JW, Morrison SS, Raphael BH, Winchell JM. Complete Genome Sequences of the Historical *Legionella pneumophila* Strains OLDA and Pontiac. Genome announcements. 2016;4(4).
106. Amaro F, Gilbert JA, Owens S, Trimble W, Shuman HA. Whole-genome sequence of the human pathogen *Legionella pneumophila* serogroup 12 strain 570-CO-H. Journal of bacteriology. 2012;194(6):1613-4.
107. Centers for Disease Control and Prevention CDC. Procedures for the Recovery of Legionella from the Environment Centers for Disease Control and Prevention (CDC) 2005:1-13.

108. Standardization IOf. Detection and Enumeration of Legionella ISO 11731. 2004.

109. Kahlisch L, Henne K, Draheim J, Brettar I, Höfle MG. High-resolution in situ genotyping of *Legionella pneumophila* populations in drinking water by multiple-locus variable-number tandem-repeat analysis using environmental DNA. Appl Environ Microbiol. 2010;76(18):6186-95.
110. Helbig JH, Bernander S, Castellani Pastoris M, Etienne J, Gaia V, Lauwers S, et al. Pan-European study on culture-proven Legionnaires' disease: distribution of *Legionella pneumophila* serogroups and monoclonal subgroups. Eur J Clin Microbiol Infect Dis. 2002;21(10):710-6.
111. Zink SD, Pedersen L, Cianciotto NP, Abu-Kwaik Y. The Dot/Icm type IV secretion system of *Legionella pneumophila* is essential for the induction of apoptosis in human macrophages. Infect Immun. 2002;70(3):1657-63.
112. Kirby JE, Vogel JP, Andrews HL, Isberg RR. Evidence for pore-forming ability by *Legionella pneumophila*. Mol Microbiol. 1998;27(2):323-36.
113. Rizzardi K, Winiecka-Krusnell J, Ramliden M, Alm E, Andersson S, Byfors S. *Legionella norrlandica* sp. nov., isolated from the biopurification systems of wood processing plants. Int J Syst Evol Microbiol. 2015;65(Pt 2):598-603.
114. Joshi AD, Swanson MS. Comparative analysis of *Legionella pneumophila* and *Legionella micdadei* virulence traits. Infect Immun. 1999;67(8):4134-42.
115. Metzker ML. Sequencing technologies - the next generation. Nature reviews Genetics. 2010;11(1):31-46.
116. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012;6(8):1621-4.
117. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature reviews Genetics. 2016;17(6):333-51.
118. Coupland P, Chandra T, Quail M, Reik W, Swerdlow H. Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. BioTechniques. 2012;53(6):365-72.
119. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics. 2012;13:341.
120. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
121. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome biology. 2014;15(11):524.
122. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-403.

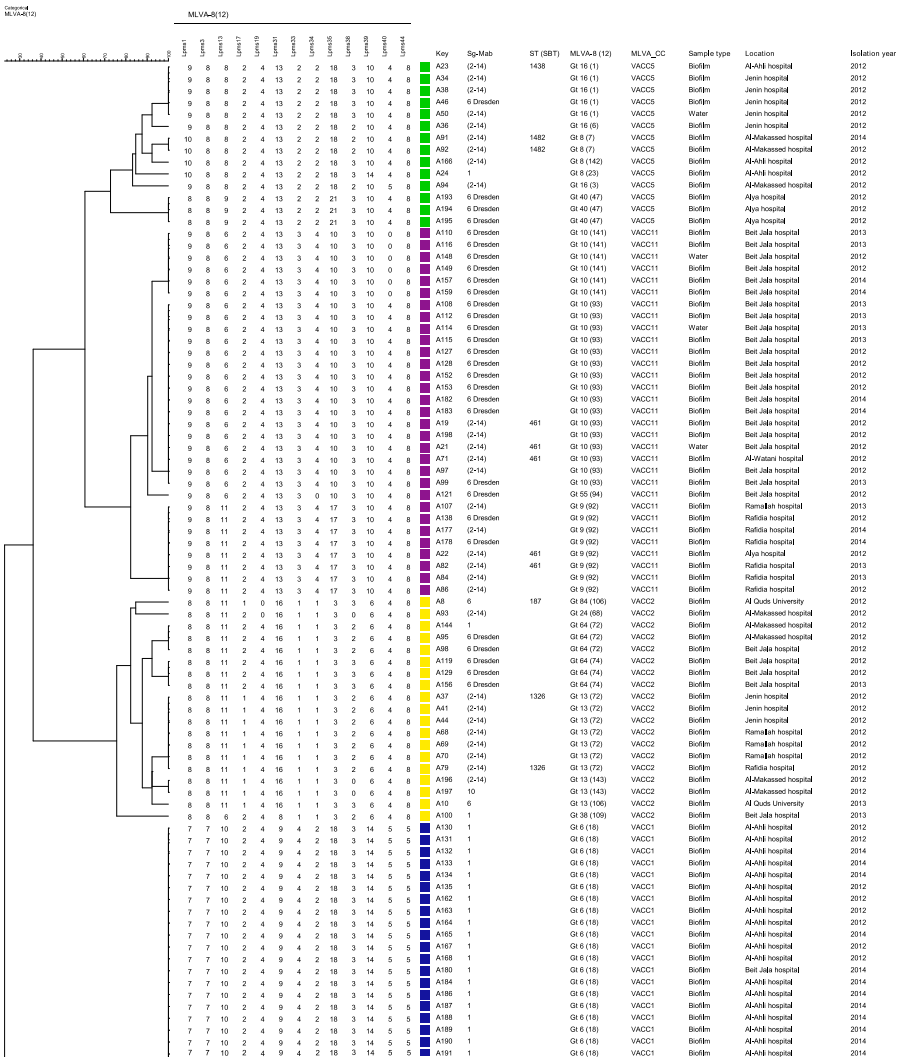
123. Darling AE, Treangen TJ, Messeguer X, Perna NT. Analyzing patterns of microbial evolution using the mauve genome alignment system. Methods in molecular biology. 2007;396:135-52.
124. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic acids research. 2017.
125. Langille MG, Hsiao WW, Brinkman FS. Evaluation of genomic island predictors using a comparative genomics approach. BMC bioinformatics. 2008;9:329.
126. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, et al. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC bioinformatics. 2006;7:142.
127. Hsiao W, Wan I, Jones SJ, Brinkman FS. IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics. 2003;19(3):418-20.
128. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic acids research. 2005;33(Database issue):D325-8.
129. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic acids research. 2016;44(D1):D694-7.
130. Clarke KR, Somerfield P. J., Gorley R. N. Testing null hypotheses in exploratory community analyses: similarity profiles and biota-environmental linkage. J Exp Mar Biol Ecol. 2008;366:56-69.
131. Petzold M, Prior K, Moran-Gilad J, Harmsen D, Lück C. Epidemiological information is key when interpreting whole genome sequence data - lessons learned from a large *Legionella pneumophila* outbreak in Warstein, Germany, 2013. Euro Surveill. 2017;22(45).
132. Segal G, Shuman HA. Characterization of a new region required for macrophage killing by *Legionella pneumophila*. Infect Immun. 1997;65(12):5057-66.
133. Berger KH, Isberg RR. Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. Mol Microbiol. 1993;7(1):7-19.
134. Segal G, Shuman HA. *Legionella pneumophila* utilizes the same genes to multiply within *Acanthamoeba castellanii* and human macrophages. Infect Immun. 1999;67(5):2117-24.
135. Cirillo SL, Bermudez LE, El-Etr SH, Duhamel GE, Cirillo JD. *Legionella pneumophila* entry gene rtxA is involved in virulence. Infect Immun. 2001;69(1):508-17.
136. Broich M, Rydzewski K, McNealy TL, Marre R, Flieger A. The global regulatory proteins LetA and RpoS control phospholipase A, lysophospholipase A, acyltransferase, and other hydrolytic activities of *Legionella pneumophila* JR32. Journal of bacteriology. 2006;188(4):1218-26.
137. Molmeret M, Bitar DM, Han L, Kwaik YA. Cell biology of the intracellular infection by *Legionella pneumophila*. Microbes and infection. 2004;6(1):129-39.
138. Wintermeyer E, Rdest U, Ludwig B, Debes A, Hacker J. Characterization of legiolysin (lly), responsible for haemolytic activity, colour production and fluorescence of *Legionella pneumophila*. Mol Microbiol. 1991;5(5):1135-43.
139. Flieger A, Rydzewski K, Banerji S, Broich M, Heuner K. Cloning and characterization of the gene encoding the major cell-associated phospholipase A of *Legionella pneumophila*, plaB, exhibiting hemolytic activity. Infect Immun. 2004;72(5):2648-58.
140. Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, et al. Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. BMC genomics. 2011;12:536.

141. Beaute J, Robesyn E, de Jong B, European Legionnaires' Disease Surveillance N. Legionnaires' disease in Europe: all quiet on the eastern front? Eur Respir J. 2013;42(6):1454-8.
142. Ma'ayeh SY, Al-Hiyasat AS, Hindiyeh MY, Khader YS. *Legionella pneumophila* contamination of a dental unit water line system in a dental teaching centre. Int J Dent Hyg. 2008;6(1):48-55.
143. Buswell CM, Herlihy YM, Lawrence LM, McGuiggan JT, Marsh PD, Keevil CW, et al. Extended survival and persistence of *Campylobacter* spp. in water and aquatic biofilms and their detection by immunofluorescent-antibody and -rRNA staining. Appl Environ Microbiol. 1998;64(2):733-41.
144. Rogers J, Dowsett AB, Dennis PJ, Lee JV, Keevil CW. Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. Appl Environ Microbiol. 1994;60(5):1585-92.
145. Williams MM, Braun-Howland EB. Growth of *Escherichia coli* in model distribution system biofilms exposed to hypochlorous acid or monochloramine. Appl Environ Microbiol. 2003;69(9):5463-71.
146. Pereira RPA, Peplies J, Höfle MG, Brettar I. Bacterial community dynamics in a cooling tower with emphasis on pathogenic bacteria and *Legionella* species using universal and genus-specific deep sequencing. Water Res. 2017;122:363-76.
147. Lasheras A, Boulestreau H, Rogues AM, Ohayon-Courtes C, Labadie JC, Gachie JP. Influence of amoebae and physical and chemical characteristics of water on presence and proliferation of *Legionella* species in hospital water systems. Am J Infect Control. 2006;34(8):520-5.
148. Zanetti F, Stampi S, De Luca G, Fateh-Moghadam P, Antonietta M, Sabattini B, et al. Water characteristics associated with the occurrence of *Legionella pneumophila* in dental units. European journal of oral sciences. 2000;108(1):22-8.
149. Shalash I GM. Hydrochemistry of the Natuf basin in Ramallah area/West Bank. Environ Geol. 2008;55:359-67.
150. Lee HK, Kang YH, Yu JY. Genomic diversity of *Legionella pneumophila* serogroup 1 from environmental water sources and clinical specimens using pulsed-field gel electrophoresis (PFGE) from 1985 to 2007, Korea. Journal of microbiology. 2010;48(5):547-53.
151. Uzel A, Ucar F, Hames-Kocabas EE. Prevalence of *Legionella pneumophila* serogroup 1 in water distribution systems in Izmir province of Turkey. APMIS. 2005;113(10):664-9.
152. Harrison TG, Afshar B, Doshi N, Fry NK, Lee JV. Distribution of *Legionella pneumophila* serogroups, monoclonal antibody subgroups and DNA sequence types in recent clinical and environmental isolates from England and Wales (2000-2008). Eur J Clin Microbiol Infect Dis. 2009;28(7):781-91.
153. Doleans A, Aurell H, Reyrolle M, Lina G, Freney J, Vandenesch F, et al. Clinical and environmental distributions of *Legionella* strains in France are different. J Clin Microbiol. 2004;42(1):458-60.
154. Martin-Loeches I, Torres A, Rinaudo M, Terraneo S, de Rosa F, Ramirez P, et al. Resistance patterns and outcomes in intensive care unit (ICU)-acquired pneumonia. Validation of European Centre for Disease Prevention and Control (ECDC) and the Centers for Disease Control and Prevention (CDC) classification of multidrug resistant organisms. J Infect. 2015;70(3):213-22.
155. Cazalet C, Jarraud S, Ghavi-Helm Y, Kunst F, Glaser P, Etienne J, et al. Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. Genome Res. 2008;18(3):431-41.

156. Kozak-Muiznieks NA, Lucas CE, Brown E, Pondo T, Taylor TH, Jr., Frace M, et al. Prevalence of sequence types among clinical and environmental isolates of *Legionella pneumophila* serogroup 1 in the United States from 1982 to 2012. J Clin Microbiol. 2014;52(1):201-11.
157. Reimer AR, Au S, Schindle S, Bernard KA. *Legionella pneumophila* monoclonal antibody subgroups and DNA sequence types isolated in Canada between 1981 and 2009: Laboratory Component of National Surveillance. Eur J Clin Microbiol Infect Dis. 2010;29(2):191-205.
158. Amemura-Maekawa J, Kikukawa K, Helbig JH, Kaneko S, Suzuki-Hashimoto A, Furuhata K, et al. Distribution of monoclonal antibody subgroups and sequence-based types among *Legionella pneumophila* serogroup 1 isolates derived from cooling tower water, bathwater, and soil in Japan. Appl Environ Microbiol. 2012;78(12):4263-70.
159. Lee HK, Shim JI, Kim HE, Yu JY, Kang YH. Distribution of *Legionella* species from environmental water sources of public facilities and genetic diversity of *L. pneumophila* serogroup 1 in South Korea. Appl Environ Microbiol. 2010;76(19):6547-54.
160. Chasqueira MJ, Rodrigues L, Nascimento M, Ramos M, Marques T. Genetic diversity and evolutionary relationships among *Legionella pneumophila* clinical isolates, Portugal, 1987 to 2012. Euro Surveill. 2014;19(46).
161. Sanchez-Buso L, Coscolla M, Pinto-Carbo M, Catalan V, Gonzalez-Candelas F. Genetic Characterization of *Legionella pneumophila* Isolated from a Common Watershed in Comunidad Valenciana, Spain. PloS one. 2013;8(4):e61564.
162. Qin T, Zhou H, Ren H, Guan H, Li M, Zhu B, et al. Distribution of sequence-based types of *legionella pneumophila* serogroup 1 strains isolated from cooling towers, hot springs, and potable water systems in China. Appl Environ Microbiol. 2014;80(7):2150-7.
163. Coscolla M, Fenollar J, Escribano I, Gonzalez-Candelas F. Legionellosis outbreak associated with asphalt paving machine, Spain, 2009. Emerg Infect Dis. 2010;16(9):1381-7.
164. Kanatani J, Isobe J, Kimata K, Shima T, Shimizu M, Kura F, et al. Molecular epidemiology of *Legionella pneumophila* serogroup 1 isolates identify a prevalent sequence type, ST505, and a distinct clonal group of clinical isolates in Toyama Prefecture, Japan. Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy. 2013;19(4):644-52.
165. Jaber L AM, Abu Tair H, Bahader S. A, Alalam H, Butmeh S, Abu Hilal D, Brettar L, Höfle M. G, Bitar D. M. In-situ sequence type analysis of *Legionella pneumophila* in respiratory tract secretions and environmental samples of a hospital in East Jerusalem. In review. 2018.
166. Kirby JE, Isberg RR. Legionnaires' disease: the pore macrophage and the legion of terror within. Trends Microbiol. 1998;6(7):256-8.
167. Sanchez-Buso L, Coscolla M, Palero F, Camaro ML, Gimeno A, Moreno P, et al. Geographical and Temporal Structures of *Legionella pneumophila* Sequence Types in Comunitat Valenciana (Spain), 1998 to 2013. Appl Environ Microbiol. 2015;81(20):7106-13.
168. Feeley JC, Gibson RJ, Gorman GW, Langford NC, Rasheed JK, Mackel DC, et al. Charcoal-yeast extract agar: primary isolation medium for *Legionella pneumophila*. J Clin Microbiol. 1979;10(4):437-41.
169. Sadosky AB, Wiater LA, Shuman HA. Identification of *Legionella pneumophila* genes required for growth within and killing of human macrophages. Infect Immun. 1993;61(12):5361-73.
170. Fernandez RC, Logan SM, Lee SH, Hoffman PS. Elevated levels of *Legionella pneumophila* stress protein Hsp60 early in infection of human monocytes and L929 cells correlate with virulence. Infect Immun. 1996;64(6):1968-76.

171. Kusnetsov J, Torvinen E, Perola O, Nousiainen T, Katila ML. Colonization of hospital water systems by *legionellae*, mycobacteria and other heterotrophic bacteria potentially hazardous to risk group patients. APMIS. 2003;111(5):546-56.
172. Moran-Gilad J, Prior K, Yakunin E, Harrison TG, Underwood A, Lazarovitch T, et al. Design and application of a core genome multilocus sequence typing scheme for investigation of Legionnaires' disease incidents. Euro Surveill. 2015;20(28).
173. Graham RM, Doyle CJ, Jennison AV. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. Epidemiol Infect. 2014;142(11):2347-51.
174. Mercante JW, Morrison SS, Desai HP, Raphael BH, Winchell JM. Genomic Analysis Reveals Novel Diversity among the 1976 Philadelphia Legionnaires' Disease Outbreak Isolates and Additional ST36 Strains. PloS one. 2016;11(9):e0164074.
175. Morozova I, Qu X, Shi S, Asamani G, Greenberg JE, Shuman HA, et al. Comparative sequence analysis of the icm/dot genes in *Legionella*. Plasmid. 2004;51(2):127-47.
176. Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. Comparative and functional genomics of *legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. Frontiers in microbiology. 2011;2:208.
177. Roy CR, Berger KH, Isberg RR. *Legionella pneumophila* DotA protein is required for early phagosome trafficking decisions that occur within minutes of bacterial uptake. Mol Microbiol. 1998;28(3):663-74.
178. Costa J, Tiago I, Da Costa MS, Verissimo A. Molecular evolution of *Legionella pneumophila* dotA gene, the contribution of natural environmental strains. Environ Microbiol. 2010;12(10):2711-29.
179. Dumenil G, Montminy TP, Tang M, Isberg RR. IcmR-regulated membrane insertion and efflux by the *Legionella pneumophila* IcmQ protein. The Journal of biological chemistry. 2004;279(6):4686-95.
180. Gomez-Valero L, Rusniok C, Rolando M, Neou M, Dervins-Ravault D, Demirtas J, et al. Comparative analyses of *Legionella* species identifies genetic features of strains causing Legionnaires' disease. Genome biology. 2014;15(11):505.
181. Bellanger X, Payot S, Leblond-Bourget N, Guedon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev. 2014;38(4):720-60.
182. Glöckner G, Albert-Weissenberger C, Weinmann E, Jacobi S, Schunder E, Steinert M, et al. Identification and characterization of a new conjugation/type IVA secretion system (trb/tra) of *Legionella pneumophila* Corby localized on two mobile genomic islands. Int J Med Microbiol. 2008;298(5-6):411-28.
183. Andersson DI, Jerlstrom-Hultqvist J, Nasvall J. Evolution of new functions de novo and from preexisting genes. Cold Spring Harbor perspectives in biology. 2015;7(6).
184. Lesnik R, Brettar I, Höfle MG. *Legionella* species diversity and dynamics from surface reservoir to tap water: from cold adaptation to thermophily. The ISME journal. 2016;10(5):1064-80.

6. Appendices



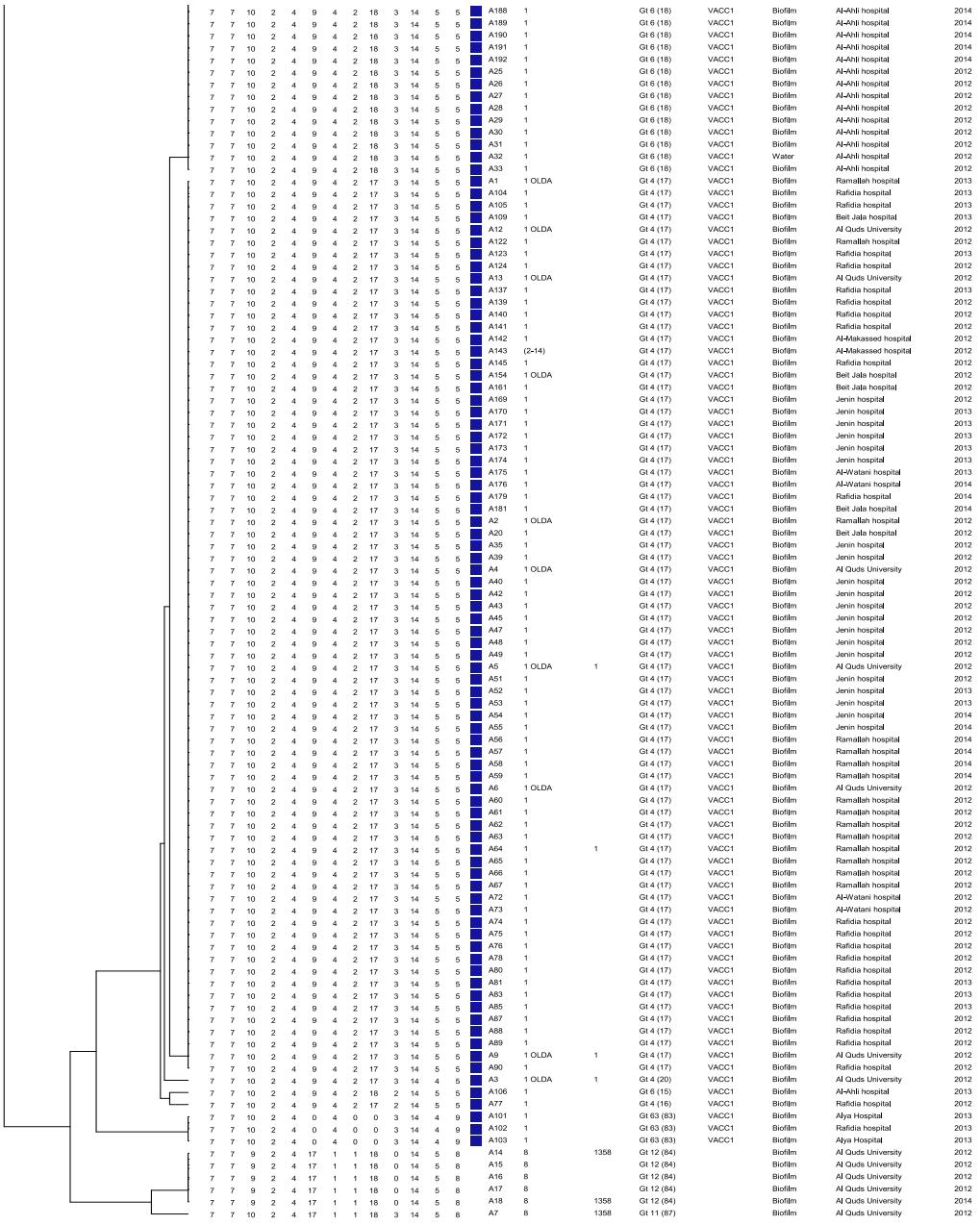


Figure A1 UPGMA inferred from the clustering analysis of the MLVA-8(12) profiles of 180 *L. pneumophila* strains isolated from water and biofilm samples taken from the Al-Quds University campus and from eight hospitals along the West Bank. MLVA clusters (VACC) of three or more genotypes were defined using a cutoff of 60% of similarity and are shown with colors done by (97).

Table A1 (Part 1): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
1st (Oct-Dec 2012)	Cold Water	HPC (CFU/L) 37°C	3.67E+04	1.00E+04	3.30E+03	1.63E+05	5.40E+04	3.90E+05	6.00E+04	3.67E+04
		HPC (CFU/L) 22°C	2.00E+04	6.70E+03	1.20E+03	1.08E+05	4.70E+04	1.90E+05	6.33E+04	3.33E+04
		Leg.count (CFU/L)	2.60E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.00E+01	0.00E+00
		Temprature °C	21.7	21.0	20.1	21.7	21.7	25.5	18.4	19.7
		pH	7.6	7.8	8.0	7.8	8.0	7.6	8.3	7.8
		Conductivity µS/cm	642	711	761	818	774	610	798	620
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		chlorine Sensitive (Cl2) mg/l	0.1	0.1	0.5	0.1	0.1	0.1	0.1	0.5
		Nitrate ($NO3^-$) mg/l	10.0	10.0	10.0	10.0	10.0	25.0	10.0	10.0
		Nitrite ($NO2^-$) mg/l	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Ammonia (NH3) mg/l	0.0	0.5	0.5	0.5	0.0	0.5	1.0	1.0
		Copper ($Cu^{+}/^{+2}$) mg/l	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Phosphate ($PO4^{3-}$) mg/l	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Zinc (Zn) mg/l	0.0	2.0	2.0	2.0	0.0	0.0	2.0	2.0
		Turb (mg/l)	1.1	1.5		1.6	1.2	0.9	0.9	
		HCO3 (mg/l)	241	176		180	210	250	202	
		Chlorides Cl (mg/l)	67.7	57.0		38.0	49.0	27.4	54.8	
		SO4 (mg/l)	15.6	12.1		14.7	20.4	24.3	35.4	
		Hardness (mg/l)	353	272		248	287	261	300	
		TDS (mg/l)	418	312		265	349	285	465	
		Flouride (mg/l)	0.2	0.3		0.2	0.2	0.2	0.2	
		Magnesium (mg/l)	33.0	39.0	38.0	40.0	35.0	28.0	34.0	36.0
		Calcium (mg/l)	102.0	78.0	79.0	84.0	82.0	87.0	91.0	90.0
		Ca/Mg ratio	3.1	2.0	2.1	2.1	2.3	3.1	2.7	2.5
		% swab positive Com	100	100	100	100	100	100	100	100
		% swab positive Lgsp	80	60	60	80	100	100	80	80
		% swab positive L1	100	60	40	80	100	100	100	80
		% swab positive Culture CIA subset	60	80	20	60	20	20	40	0
		% positive Lpn PCR water	100	50	0	50	0	100	100	0
		% Total positive CIA (*L.pneumophila*) (W+B)	100.0	57.1	33.3	71.4	71.4	100.0	100.0	66.7
		% positive swabs (culture Based)	47.7	38.6	6.8	38.6	13.6	4.5	22.7	2.3
		% VACC1	50.0	40.6	40.0	77.8	0.0	3.1	26.5	0.0
		% VACC2	10.7	3.1	0.0	16.7	20.0	0.0	0.0	0.0
		% VACC5	17.9	0.0	0.0	0.0	30.0	0.0	5.9	0.0
		% VACC11	0.0	9.4	20.0	0.0	0.0	6.3	0.0	16.7
		% Gt4(17)	50.0	37.5	40.0	77.8	0.0	3.1	0.0	0.0
		% Gt6(18)	0.0	0.0	0.0	0.0	0.0	0.0	26.5	0.0
		% Gt10(93)	0.0	0.0	20.0	0.0	0.0	6.3	0.0	0.0
		% Gt13(72)	10.7	3.1	0.0	16.7	0.0	0.0	0.0	0.0
		% Gt9(92)	0.0	9.4	0.0	0.0	0.0	0.0	0.0	16.7
		% Gt10(141)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt12(84)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt16(1)	14.3	0.0	0.0	0.0	0.0	0.0	2.9	0.0
		% Gt40(47)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt63(83)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt64(74)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Hot Water	HPC (CFU/l) 37°C	NA	0.00E+00	NA	1.17E+05	5.67E+04	2.78E+05	3.67E+04	NA
		HPC (CFU/l) 22°C	NA	0.00E+00	NA	3.00E+04	6.70E+03	1.95E+05	4.67E+04	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	0.00E+00	0.00E+00	3.50E+01	0.00E+00	NA
		Temprature °C	NA	70.9	NA	44.5	51.2	52.2	38.4	NA
		pH	NA	7.6	NA	7.6	8.0	7.6	8.2	NA
		Conductivity µS/cm	NA	701	NA	815	785	620	802	NA

Table A1 (Part 2): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
2nd (Mar-May 2013)	Cold Water	HPC (CFU/L) 37°C	1.40E+03	1.60E+03	2.56E+02	7.33E+04	1.50E+03	2.50E+04	1.53E+04	2.50E+03
		HPC (CFU/L) 22°C	1.30E+03	1.67E+03	1.93E+02	3.33E+04	8.30E+02	2.30E+04	3.30E+03	2.00E+03
		Leg.count (CFU/L)	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
		Temprature °C	25	23.2	25.5	22.5	24	26.2	24.2	21.3
		pH	7.6	7.8	8	7.6	7.82	7.8	7.6	7.8
		Conductivity µS/cm	650	720	780	650	761	650	780	998
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2	2	2	2	2	2	2	2
		chlorine Sensitive (Cl2) mg/l	5	0.5	0.5	0.1	0.1	0.5	0.1	0.1
		Nitrate ($NO3^-$) mg/l	10	25	25	10	10	10	10	10
		Nitrite ($NO2^-$) mg/l	0	0	0	0	0	0	0	0
		Ammonia (NH3) mg/l	0.5	0.5	0.5	0.5	0	0.5	0.5	0.5
		Copper ($Cu^{+}/^{+2}$) mg/l	0	0	0	0	0	0	0	0
		Phosphate ($PO4^{3-}$) mg/l	0	0	0	0	0	0	0	0
		Zinc (Zn) mg/l	2	2	2	2	0	2	2	2
		Turb (mg/l)	1.2	1.6		1.8	1.4	1.1	1.1	
		HCO3 (mg/l)	242	180		178	206	234	197	
		Chlorides CL (mg/l)	78	55		36	50	25	58	
		SO4 (mg/l)	21	14		15	20	19	30	
		Hardness (mg/l)	312	296		228	274	263	273	
		TDS (mg/l)	371	320		283	334	285	410	
		Flouride (mg/l)	0.2	0.2		0.2	0.2	0.2	0.2	
		Magnesium (mg/l)	34.0	39.5	38.0	39.5	35.5	25.0	32.0	33.5
		Calcium (mg/l)	99.5	81.0	82.0	82.5	81.5	83.5	92.5	92.0
		Ca/Mg ratio	2.9	2.1	2.2	2.1	2.3	3.3	2.9	2.7
		% swab positive Com	100	100	100	100	100	100	100	100
		% swab positive Lgsp	80	80	20	60	40	80	60	60
		% swab positive L1	60	60	20	60	40	80	60	40
		% swab positive Culture CIA subset	0	60	0	20	0	60	20	40
		% positive Lpn PCR water	100	0	0	0	0	100	100	0
		% Total positive CIA (*L.pneumophila)* (W+B)	66.7	42.9	16.7	42.9	28.6	85.7	71.4	33.3
		% positive swabs (culture Based)	0	20	0	0	0	20	5	10
		% VACC1	0.0	9.4	0.0	0.0	0.0	0.0	2.9	33.3
		% VACC2	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0
		% VACC5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% VACC11	0.0	3.1	0.0	0.0	0.0	6.3	0.0	0.0
		% Gt4(17)	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt6(18)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt10(93)	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0
		% Gt13(72)	0.0	0.0	0.0	0.0	0.0	3.1	0.0	0.0
		% Gt9(92)	0.0	3.1	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt10(141)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt12(84)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt16(1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt40(47)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		% Gt63(83)	0.0	3.1	0.0	0.0	0.0	0.0	0.0	33.3
		% Gt64(74)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Hot Water	HPC (CFU/L) 37°C	NA	0.00E+00	NA	NA	7.60E+02	9.34E+03	2.70E+04	NA
		HPC (CFU/L) 22°C	NA	0.00E+00	NA	NA	6.33E+02	8.30E+03	4.30E+04	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	NA	0.00E+00	0.00E+00	0.00E+00	NA
		Temprature °C	NA	70.9	NA	NA	47.4	44.8	42.7	NA
		pH	NA	7.6	NA	NA	7.9	7.8	7.6	NA
		Conductivity µS/cm	NA	701	NA	NA	781	630	799	NA

Table A1 (Part 3): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
3rd (Jun-Aug 2013)	Cold Water	HPC (CFU/L) 37°C	3.70E+04	4.60E+04	8.50E+03	2.47E+05	1.50E+04	1.50E+05	3.70E+04	4.50E+04
		HPC (CFU/L) 22°C	1.00E+04	2.13E+04	8.00E+03	2.11E+05	8.30E+03	1.03E+05	3.00E+04	2.00E+04
		Leg.count (CFU/L)	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.67E+02	0.00E+00	0.00E+00
		Temprature °C	26.6	25.5	26.2	25.3	33.6	27	22	23.1
		pH	8	7.9	8	7.6	7.6	7.6	7.6	7.8
		Conductivity µS/cm	628	720	743	650	734	620	780	898
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2	2	2	2	2	2	2	2
		chlorine Sensitive (Cl2) mg/l	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
		Nitrate ($NO3^-$) mg/l	10	10	25	10	0	10	10	10
		Nitrite ($NO2^-$) mg/l	0	0	0	0	0	0	0	0
		Ammonia (NH3) mg/l	0.5	0.5	0.5	0.5	0.5	0	0.5	0.5
		Copper ($Cu^{+}/^{+2}$) mg/l	0	0	0	0	0	0	0	0
		Phosphate ($PO4^{3-}$) mg/l	0	0	0	0	0	0	0	0
		Zinc (Zn) mg/l	0	0	0	0	0	0	0	0
		Turb (mg/l)	1.3	1.8		2.0	1.6	1.4	1.3	
		HCO3 (mg/l)	243	183		175	202	217	192	
		Chlorides CL (mg/l)	89	53		34	52	23	60	
		SO4 (mg/l)	26	16		16	19	14	26	
		Hardness (mg/l)	270	320		207	261	264	246	
		TDS (mg/l)	325	329		301	319	285	356	
		Flouride (mg/l)	NA	0.17		NA	0.2	0.17	0.2	
		Magnesium (mg/l)	35	40	38	39	36	22	30	31
		Calcium (mg/l)	96	84	85	81	81	80	94	94
		Ca/Mg ratio	2.74	2.10	2.24	2.08	2.25	3.64	3.13	3.03
		% swab positive Com	100	100	100	100	100	100	100	100
		% swab positive Lgsp	80	60	40	60	40	100	100	80
		% swab positive L1	80	60	20	60	20	100	100	60
		% swab positive Culture CIA subset	0	20	0	20	0	100	0	0
		% positive Lpn PCR water	100	50	0	50	0	100	100	0
		% Total positive CIA (*L.pneumophila*) (W+B)	83.3	57.1	16.7	57.1	14.3	100.0	100.0	50.0
		% positive swabs (culture Based)	10	10	0	5	0	60	0	0
		% VACC1	0.00	6.25	0.00	5.56	0.00	3.13	0.00	0.00
		% VACC2	0.00	0.00	0.00	0.00	0.00	3.13	0.00	0.00
		% VACC5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% VACC11	0.00	0.00	0.00	0.00	0.00	21.88	0.00	0.00
		% Gt4(17)	0.00	6.25	0.00	5.56	0.00	3.13	0.00	0.00
		% Gt6(18)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt10(93)	0.00	0.00	0.00	0.00	0.00	12.50	0.00	0.00
		% Gt13(72)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt9(92)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt10(141)	0.00	0.00	0.00	0.00	0.00	6.25	0.00	0.00
		% Gt12(84)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt16(1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt40(47)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt63(83)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		% Gt64(74)	0.00	0.00	0.00	0.00	0.00	3.13	0.00	0.00
	Hot Water	HPC (CFU/L) 37°C	NA	0.00E+00	NA	6.08E+05	7.70E+03	6.56E+05	2.70E+04	NA
		HPC (CFU/L) 22°C	NA	0.00E+00	NA	6.10E+05	6.33E+03	2.00E+05	4.30E+04	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	0.00E+00	0.00E+00	5.08E+02	0.00E+00	NA
		Temprature °C	NA	55.6	NA	43.2	49.7	40.3	42.7	NA
		pH	NA	8	NA	7.7	7.6	7.8	7.6	NA
		Conductivity µS/cm	NA	758	NA	678	786	697	799	NA

Table A1 (Part 4): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
4th (Mar-May 2014)	Cold Water	HPC (CFU/L) 37°C	6.60E+03	5.00E+03	6.50E+03	NA	1.60E+03	5.33E+03	8.33E+03	1.35E+04
		HPC (CFU/L) 22°C	3.30E+03	2.33E+03	4.00E+03	NA	1.00E+03	1.10E+03	6.60E+03	1.60E+03
		Leg.count (CFU/L)	0.00E+00	0.00E+00	0.00E+00	NA	0.00E+00	0.00E+00	0.00E+00	0.00E+00
		Temprature °C	21.6	26.2	21.4	NA	23.1	21.2	19.1	20.5
		pH	7.7	8.2	7.97	NA	8	7.98	8.1	7.9
		Conductivity µS/cm	860	560	653	NA	595	421	447	578
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2	0	2	NA	0	2	0	0
		chlorine Sensitive (Cl2) mg/l	0.5	0.5	0.5	NA	0.1	0.5	10	0.5
		Nitrate ($NO3^-$) mg/l	10	0	0	NA	0	10	10	10
		Nitrite ($NO2^-$) mg/l	0	0	0	NA	0	0	0	0
		Ammonia (NH3) mg/l	0.5	0.5	0.5	NA	0	0.5	0	0
		Copper ($Cu^{+}/^{+2}$) mg/l	0	0	0	NA	0	0	0	0
		Phosphate ($PO4^{3-}$) mg/l	0	0	0	NA	0	0	0	0
		Zinc (Zn) mg/l	0	0	0	NA	0	0	2	2
		Turb (mg/l)	1.3	1.6		NA	1.4	1.4	1.0	
		HCO3 (mg/l)	230	181		NA	197	219	198	
		Chlorides CL (mg/l)	99	51		NA	56	26	63	
		SO4 (mg/l)	25	14		NA	20	12	32	
		Hardness (mg/l)	269	308		NA	260	260	254	
		TDS (mg/l)	358	327		NA	325	283	371	
		Flouride (mg/l)	0.2	0.2		NA	0.2	0.2	0.2	
		Magnesium (mg/l)	34.0	38.5	38.0	NA	33.5	21.0	29.0	31.5
		Calcium (mg/l)	96.5	84.5	83.5	NA	82.0	78.5	91.5	93.5
		Ca/Mg ratio	2.8	2.2	2.2	NA	2.4	3.7	3.2	3.0
		% swab positive Com	100	100	100	NA	100	100	100	100
		% swab positive Lgsp	60	80	40	NA	80	80	100	80
		% swab positive L1	40	80	20	NA	80	80	100	60
		% swab positive Culture CIA subset	0	80	0	NA	60	60	100	0
		% positive Lpn PCR water	100	0	0	NA	0	100	100	0
		% Total positive CIA (*L.pneumophila*) (W+B)	50.0	57.1	16.7	NA	57.1	85.7	100.0	50.0
		% positive swabs (culture Based)	0	30	0	NA	15	15	30	0
		% VACC1	0.0	12.5	0.0	NA	20.0	0.0	17.6	0.0
		% VACC2	0.0	0.0	0.0	NA	10.0	3.1	0.0	0.0
		% VACC5	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% VACC11	0.0	3.1	0.0	NA	0.0	6.3	0.0	0.0
		% Gt4(17)	0.0	12.5	0.0	NA	0.0	0.0	0.0	0.0
		% Gt6(18)	0.0	0.0	0.0	NA	0.0	0.0	17.6	0.0
		% Gt10(93)	0.0	0.0	0.0	NA	0.0	6.3	0.0	0.0
		% Gt13(72)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt9(92)	0.0	3.1	0.0	NA	0.0	0.0	0.0	0.0
		% Gt10(141)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt12(84)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt16(1)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt40(47)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt63(83)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt64(74)	0.0	0.0	0.0	NA	0.0	3.1	0.0	0.0
	Hot Water	HPC (CFU/l) 37C	NA	4.70E+04	NA	NA	8.70E+04	1.30E+04	1.00E+03	NA
		HPC (CFU/l) 22C	NA	1.73E+04	NA	NA	1.16E+03	1.30E+03	3.16E+03	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	NA	0.00E+00	0.00E+00	0.00E+00	NA
		Temprature °C	NA	38.8	NA	NA	39.8	41.8	47.6	NA
		pH	NA	8.0	NA	NA	7.9	7.9	8.1	NA
		Conductivity µS/cm	NA	680	NA	NA	754	473	367	NA

Table A1 (Part 5): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
5th (Jun-Aug 2014)	Cold Water	HPC (CFU/l) 37°C	3.00E+04	5.03E+04	5.40E+04	NA	2.00E+04	2.07E+04	4.00E+04	6.30E+04
		HPC (CFU/l) 22°C	1.60E+03	3.76E+04	5.30E+04	NA	6.60E+02	1.83E+04	1.90E+04	2.20E+04
		Leg.count (CFU/L)	0.00E+00	0.00E+00	0.00E+00	NA	0.00E+00	4.21E+02	0.00E+00	0.00E+00
		Temprature °C	23.6	26.2	21.4	NA	29.1	27.5	25	20.1
		pH	7.7	8.2	8.0	NA	8.4	7.7	8.1	7.7
		Conductivity µS/cm	984	560	653	NA	737	615	681	607
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2	0	2	NA	0	2	0	0
		chlorine Sensitive (Cl2) mg/l	1	0.5	0.5	NA	0.1	0.5	1	0.5
		Nitrate ($NO3^-$) mg/l	10	0	0	NA	0	10	10	10
		Nitrite ($NO2^-$) mg/l	0	0	0	NA	0	0	0	0
		Ammonia (NH3) mg/l	0.5	0.5	0.5	NA	0	0.5	0	0
		Copper ($Cu^{+}/^{+2}$) mg/l	0	0	0	NA	0	0	0	0
		Phosphate ($PO4^{3-}$) mg/l	0	0	0	NA	0	0	0	0
		Zinc (Zn) mg/l	0	0	0	NA	0	0	2	0
		Turb (mg/l)	1.3	1.4		NA	1.2	1.3	0.7	
		HCO3 (mg/l)	217.4	179.4		NA	192.7	220.6	205.0	
		Chlorides CL (mg/l)	108.0	48.2		NA	59.3	29.6	66.0	
		SO4 (mg/l)	24.3	12.8		NA	20.2	11.3	37.5	
		Hardness (mg/l)	266.8	297.0		NA	258.6	256.5	262.0	
		TDS (mg/l)	390.7	325.0		NA	331.8	281.2	387.0	
		Flouride (mg/l)	0.2	0.2		NA	0.2	0.2	0.2	
		Magnesium (mg/l)	34.0	38.0	38.0	NA	33.0	21.0	29.0	32.0
		Calcium (mg/l)	94.0	85.0	82.0	NA	83.0	78.0	91.0	95.0
		Ca/Mg ratio	2.8	2.2	2.2	NA	2.5	3.7	3.1	3.0
		% swab positive Com	100	100	100	NA	100	100	100	100
		% swab positive Lgsp	80	80	80	NA	60	100	100	80
		% swab positive L1	60	60	80	NA	60	100	100	80
		% swab positive Culture CIA subset	0	20	40	NA	0	60	40	0
		% positive Lpn PCR water	0	100	100	NA	0	100	100	100
		% Total positive CIA (*L.pneumophila*) (W+B)	50	71.42857	83.333333	NA	42.857143	100	100	83.3333333
		% positive swabs (culture Based)	0	15	0	NA	0	65	35	0
		% VACC1	0.0	3.1	0.0	NA	0.0	6.3	17.6	0.0
		% VACC2	0.0	0.0	0.0	NA	0.0	3.1	0.0	0.0
		% VACC5	0.0	0.0	0.0	NA	0.0	0.0	2.9	0.0
		% VACC11	0.0	0.0	0.0	NA	0.0	18.8	0.0	0.0
		% Gt4(17)	0.0	3.1	0.0	NA	0.0	6.3	0.0	0.0
		% Gt6(18)	0.0	0.0	0.0	NA	0.0	0.0	17.6	0.0
		% Gt10(93)	0.0	0.0	0.0	NA	0.0	6.3	0.0	0.0
		% Gt13(72)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt9(92)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt10(141)	0.0	0.0	0.0	NA	0.0	12.5	0.0	0.0
		% Gt12(84)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt16(1)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt40(47)	0.00	0.00	0.00	NA	0.00	0.0	0.00	0.00
		% Gt63(83)	0.00	0.00	0.00	NA	0.00	0.0	0.00	0.00
		% Gt64(74)	0.00	0.00	0.00	NA	0.00	3.1	0.00	0.00
	Hot Water	HPC (CFU/L) 37°C	NA	6.70E+04	NA	NA	4.40E+03	3.70E+04	2.47E+04	NA
		HPC (CFU/L) 22°C	NA	3.37E+04	NA	NA	2.00E+03	4.60E+03	2.80E+04	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	NA	0.00E+00	0.00E+00	0.00E+00	NA
		Temprature °C	NA	38.8	NA	NA	54.2	37.3	33.7	NA
		pH	NA	8.0	NA	NA	7.6	8.1	8.1	NA
		Conductivity µS/cm	NA	680	NA	NA	708	615	661	NA

Table A1 (Part 6): Detailed data collection on sampling per site and sampling campaign

Collection No.	Water Type	Physiochem parameter	Hospital A (Jenin)	Hospital B (Nablus)	Hospital C (Nablus)	Hospital D (Ramallah)	Hospital E (Jerusalem)	Hospital F (Bethlehem)	Hospital G (Hebron)	Hospital H (Hebron)
6th (Oct-Dec 2014)	Cold Water	HPC (CFU/L) 37°C	1.33E+04	1.00E+04	1.67E+04	NA	1.00E+04	4.00E+04	1.33E+04	3.30E+03
		HPC (CFU/L) 22°C	8.40E+03	1.67E+04	1.20E+04	NA	3.30E+04	4.40E+04	1.10E+04	2.70E+03
		Leg.count (CFU/L)	0.00E+00	0.00E+00	0.00E+00	NA	0.00E+00	0.00E+00	0.00E+00	0.00E+00
		Temprature °C	20.3	23.8	23.7	NA	19.8	20.8	30.7	29.9
		pH	7.6	8.2	8.0	NA	7.8	8.0	8.2	8.1
		Conductivity µS/cm	913	560	653	NA	595	421	400	542
		Total iron (Fe^{+3}/Fe^{+2}) mg/l	2	0	2	NA	0	2	0	0
		chlorine Sensitive (Cl2) mg/l	0.5	0.5	0.5	NA	0.1	0.5	10	0.5
		Nitrate ($NO3^-$) mg/l	0	0	0	NA	0	10	10	10
		Nitrite ($NO2^-$) mg/l	0	0	0	NA	0	0	0	0
		Ammonia (NH3) mg/l	0	0.5	0.5	NA	0	0.5	0	0
		Copper ($Cu^{+}/^{+2}$) mg/l	0	0	0	NA	0	0	0	0
		Phosphate ($PO4^{3-}$) mg/l	0	0	0	NA	0	0	0	0
		Zinc (Zn) mg/l	0	0	0	NA	0	0	2	2
		Turb (mg/l)	1.3	1.5		NA	1.3	1.3	0.9	
		HCO3 (mg/l)	224	180		NA	206	220	202	
		Chlorides CL (mg/l)	103	49		NA	61	28	65	
		SO4 (mg/l)	25	14		NA	21	12	35	
		Hardness (mg/l)	268	303		NA	272	258	258	
		TDS (mg/l)	374	326		NA	340	282	379	
		Flouride (mg/l)	0.2	0.2		NA	0.2	0.2	0.2	
		Magnesium (mg/l)	31	37	38	NA	34	29	30	32
		Calcium (mg/l)	88	79	82	NA	75	84	96	95
		Ca/Mg ratio	2.84	2.14	2.16	NA	2.21	2.90	3.20	2.97
		swab % positive Com	100	100	100	NA	100	100	100	100
		swab % positive Lgsp	100	80	60	NA	60	100	100	60
		swab % positive L1	100	60	60	NA	60	80	100	40
		swab% positive Culture CIA subset	100	60	40	NA	40	60	100	40
		% positive Lpn PCR water	0	0	0	NA	0	100	100	0
		% Total positive CIA (*L.pneumophila*) (W+B)	83.3	42.9	50.0	NA	42.9	85.7	100.0	33.3
		% positive swabs (culture Based)	30	15	10	NA	10	15	50	15
		% VACC1	21.4	3.1	40.0	NA	0.0	3.1	26.5	0.0
		% VACC2	0.0	0.0	0.0	NA	20.0	0.0	0.0	0.0
		% VACC5	0.0	0.0	0.0	NA	0.0	0.0	0.0	50.0
		% VACC11	0.0	6.3	0.0	NA	0.0	9.4	0.0	0.0
		% Gt4(17)	21.4	3.1	40.0	NA	0.0	3.1	0.0	0.0
		% Gt6(18)	0.0	0.0	0.0	NA	0.0	0.0	26.5	0.0
		% Gt10(93)	0.0	0.0	0.0	NA	0.0	9.4	0.0	0.0
		% Gt13(72)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt9(92)	0.0	6.3	0.0	NA	0.0	0.0	0.0	0.0
		% Gt10(141)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt12(84)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt16(1)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt40(47)	0.0	0.0	0.0	NA	0.0	0.0	0.0	50.0
		% Gt63(83)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
		% Gt64(74)	0.0	0.0	0.0	NA	0.0	0.0	0.0	0.0
	Hot Water	HPC (CFU/L) 37°C	NA	1.67E+04	NA	NA	2.06E+03	3.70E+04	3.33E+04	NA
		HPC (CFU/L) 22°C	NA	1.33E+04	NA	NA	3.30E+03	4.60E+03	1.47E+04	NA
		Leg.count (CFU/L)	NA	0.00E+00	NA	NA	0.00E+00	0.00E+00	0.00E+00	NA
		Temprature °C	NA	36.3	NA	NA	44.8	44.3	30.7	NA
		pH	NA	8.0	NA	NA	8	7.9	8.5	NA
		Conductivity µS/cm	NA	680	NA	NA	754	672	400	NA

Table A2: Genotype niches summary matrix gentype by genotype.

Parameter	Genotype	Gt4(17)	Gt6(18)	Gt10(93)	Gt13(72)	Gt9(92)	Gt10(141)	Gt16(1)	Gt40(47)	Gt64(74)	Gt63(83)
Leg.count (CFU/L) in water	Gt4(17)										
	Gt6(18)	NS									
	Gt10(93)	*	*								
	Gt13(72)	NS	NS	NS							
	Gt9(92)	***	***	**	NS						
	Gt10(141)	***	***	**	***	***					
	Gt16(1)	***	***	NS	**	***	***				
	Gt40(47)	***	***	**	NS	***	***	***			
	Gt64(74)	**	NS	NS	NS	NS	NS	NS	NS		
	Gt63(83)	***	***	**	NS	***	***	***	NS	NS	
Turb (mg/l)	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	*	***								
	Gt13(72)	NS	***	NS							
	Gt9(92)	NS	***	*	NS						
	Gt10(141)	*	***	NS	NS	NS					
	Gt16(1)	***	***	*	**	**	***				
	Gt40(47)	***	NS	**	***	**	***	*			
	Gt64(74)	NS	***	NS	NS	NS	NS	**	**		
	Gt63(83)	NS	NS	NS	NS	NS	NS	NS	NS	NS	
Chloride Cl (mg/l)	Gt4(17)										
	Gt6(18)	*									
	Gt10(93)	***	***								
	Gt13(72)	NS	NS	**							
	Gt9(92)	NS	**	***	NS						
	Gt10(141)	**	***	NS	**	***					
	Gt16(1)	NS	*	***	NS	***	***				
	Gt40(47)	***	**	***	NS	***	***	NS			
	Gt64(74)	*	***	NS	*	***	NS	***	***		
	Gt63(83)	NS	NS	***	NS	NS	***	NS	NS	**	
SO4 (mg/l)	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	NS	***								
	Gt13(72)	NS	***	*							
	Gt9(92)	NS	***	NS	NS						
	Gt10(141)	**	***	NS	***	*					
	Gt16(1)	*	***	**	NS	NS	***				
	Gt40(47)	***	**	NS	**	*	***	NS			
	Gt64(74)	NS	***	**	**	NS	NS	*	*		
	Gt63(83)	NS	**	NS	NS	NS	**	NS	NS	NS	
TDS (mg/l)	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	**	***								
	Gt13(72)	NS	*	NS							
	Gt9(92)	NS	**	***	NS						
	Gt10(141)	***	***	*	NS	**					
	Gt16(1)	***	NS	***	**	**	***				
	Gt40(47)	***	**	***	NS	NS	***	*			
	Gt64(74)	NS	***	NS	NS	NS	NS	***	**		
	Gt63(83)	NS	NS	NS	NS	NS	*	*	NS	*	
Mg (mg/l)	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	***	***								
	Gt13(72)	NS	*	***							
	Gt9(92)	*	***	***	NS						
	Gt10(141)	***	***	NS	***	***					
	Gt16(1)	*	**	***	**	***	***				
	Gt40(47)	***	NS	*	*	***	***	*			
	Gt64(74)	**	***	NS	**	***	NS	***	**		
	Gt63(83)	NS	**	***	NS	NS	*	NS	NS	**	
Ca (mg/l)	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	*	***								
	Gt13(72)	NS	NS	NS							
	Gt9(92)	NS	***	NS	*						
	Gt10(141)	*	***	*	**	NS					
	Gt16(1)	**	*	***	*	***	***				
	Gt40(47)	***	***	***	NS	***	***	NS			
	Gt64(74)	NS	***	NS	NS	NS	NS	***	***		
	Gt63(83)	NS	NS	*	NS	*	NS	NS	NS	NS	
Ca/Mg ratio	Gt4(17)										
	Gt6(18)	***									
	Gt10(93)	***	*								
	Gt13(72)	NS	*	***							
	Gt9(92)	NS	***	***	*						
	Gt10(141)	***	***	**	***	***					
	Gt16(1)	***	NS	*	NS	***	***				
	Gt40(47)	***	***	***	NS	***	***	NS			
	Gt64(74)	***	***	NS	**	***	NS	***	*		
	Gt63(83)	NS	**	*	NS	NS	***	*	NS	**	

Legend: Independent t-test *: *P<0.05* , **:*P<0.01* , ***:*P<0.001* , NA: Not Avaliable and NS: Non significant. Group A: Green, group B: Red and group C: Blue. Significant marks highlighted light green: for genotypes from different groups and red: for genotypes

Table A3: Main features of *L. pneumophila* isolates and reference strains used in the study.

Strain designation	Reference No:	Sg (mAb)[1]	ST[2]	MLVA-8(12)[3]	VACC[3]	Site of Isolation	Year of Isolation	Source of isolation	Genome Size (bp)	No. of Genes	Genomic Islands (bp)	Genomic Islands (%)	CDS*	Contigs	Reason for selection
Lpn-LPE509	CP003886	NA	NA	NA	NA	Shanghai, China	2013	Hospital water	3,434,224	3,105	NA	NA	3,055	1	Already published
Lpn-Philadelphia 1	NC_002942	1 Philadelphia	ST36	Gt 64 (74)	VACC2	Philadelphia, USA	1974	Clinical sample	3,397,754	2,943	180,555	5.3	3,023	1	Type strain- well characterized/ cause Philadelphia outbreak/ Already published
Lpn-Thunderbay	CP003730	Sg.6	ST187	NA	NA	Thunderbay, Canada	Unknown	Clinical sample	3,455,167	2,998	NA	NA	3,116	1	Already published
Lpn-ATCC43290	CP003192.1	Sg.12	ST187	NA	NA	USA	1987	Clinical sample	3,359,001	2,926	NA	NA	2,993	1	Already published
Lpn-lpm7613	NZ_LT598657, NZ_LT598658	NA	NA	NA	NA	NA	2000	Clinical sample	3,261,562	2,944	NA	NA	2,894	1	Already published
Lpn-Lens	GCA_000048665.1	Sg.1	ST15	NA	NA	France	2004	Clinical sample	3,345,687	3,004	180,986	5.4	2,956	1	Caused a big outbreak in France/ already published
Lpn-Lorraine	FQ958210	Sg.1	ST47	NA	NA	France	2004	Clinical sample	3,467,254	3,069	0	0	3,130	1	In top six strains that cause disease/ already published
Lpn-HL06041035	FQ958211.1	NA	NA	NA	NA	France	2006	Hospital water	3,492,535	3,184	NA	NA	3,113	1	Already published
Lpn-D7631	CP015343	Sg.1	ST731	NA	NA	NYC, USA	2012	Environmental sample	3,436,178	3,090	NA	NA	3,040	1	Already published
Lpn-D7630	CP015344	Sg.1	ST731	NA	NA	NYC, USA	2012	Clinical sample	3,444,702	3,097	NA	NA	3,047	1	Already published

Lpn-D7632	CP015342	Sg.1	ST731	NA	NA	NYC, USA	2012	Clinical sample	3,435,648	3,087	NA	NA	3,037	1	Already published
Lpn-Paris	CIP107629	1 Philadelphia	ST1	Gt 4 (17)	VACC1	Paris, France	2004	Clinical sample	3,503,610	3,224	163,637	4.7	3,101	1	Worldwide distributed/ already published/ manually annonated
Lpn-OLDA	CP016030	1 OLDA	ST1	NA	NA	USA	1947	Frozen cell culture	3,486,108	3,144	0	0	3,094	1	already published
Lpn-Pontiac	CP016029	1 OLDA	NA	NA	NA	Michigan, USA	1968	Sporadic LD case	3,545,001	3,224	NA	NA	3,174	1	Already published
Lpn-Toronto	CP012019.1	NA	NA	NA	NA	Toronto, Canada	2005	Clinical sample	3,573,898	3,269	NA	NA	3,187	1	Already published
Lpn-Alcoy 2300/99	CP001828	Sg.1	ST578	NA	NA	Alcoy, Spain	1999	Clinical sample	3,516,334	3,190	157,442	4.5	NA	1	Caused big outbreak in Spain/ already published/ manually annotated
Lpn-Corby	CP000675	1 Knoxville	ST1685 or ST51	Gt 86 (97)	NA	UK	1985	Clinical sample	3,576,470	3,204	217,089	6.1	NA	1	already published/manually annotated
A156_Gt64(74)_Ps		6 Dresden	ST74	Gt 64 (74)	VACC2	Beit Jala, Palestine	2013	Hopital environmental swab	3,366,323	3,049	174,331	5.2	3008	40	Thunderbay branch
A129_Gt64(74)_Ps		6 Dresden	ST74	Gt 64 (74)	VACC2	Beit Jala, Palestine	2012	Hopital environmental swab	3,527,232	3,206	293,904	8.3	3166	56	Thunderbay branch
A15_Gt12(84)_Ps		Sg.8	ST 1358	Gt 12 (84)	NA	Jerusalem, Palestine	2012	Hopital environmental swab	3,495,260	3,088	180,013	5.2	3,053	127	Lorraine brach

A139_Gt4(17)_Ps		Sg.1	ST1	Gt 4 (17)	VACC1	Nablus, Palestine	2012	Hopital environmental swab	3,579,958	3,242	223,851	6.3	3,205	78	Worldwide present and highest abundance in our collections
A5_Gt4(17)_Ps		1 OLDA	ST1	Gt 4 (17)	VACC1	Jerusalem, Palestine	2012	Hopital environmental swab	3,586,669	3,249	225,997	6.3	3,211	80	Worldwide present and highest abundance in our collections
A29_Gt6(18)_Ps		Sg.1	ST1	Gt 6 (18)	VACC1	Hebron, Palestine	2012	Hopital environmental swab	3,626,550	3,289	339,005	9.3	3,251	68	Dominant in southern West Bank
A131_Gt6(18)_Ps		Sg.1	ST1	Gt 6 (18)	VACC1	Hebron, Palestine	2012	Hopital environmental swab	3,548,746	3,206	281,219	7.9	3,169	67	Dominant in southern West Bank
A193_Gt40(47)_Ps		6 Dresden	ST345	Gt 40 (47)	VACC5	Hebron, Palestine	2014	Hopital environmental swab	3,446,379	3,098	185,152	5.4	3,056	33	Alcoy branch, close to Warstein outbreak
A194_Gt40(47)_Ps		6 Dresden	ST345	Gt 40 (47)	VACC5	Hebron, Palestine	2014	Hopital environmental swab	3,467,907	3,121	156,728	4.5	3,071	1	Alcoy branch, close to Warstein outbreak
A195_Gt40(47)_Ps		6 Dresden	ST345	Gt 40 (47)	VACC5	Hebron, Palestine	2014	Hopital environmental swab	3,440,802	3,095	167,241	4.9	3,051	42	Alcoy branch, close to Warstein outbreak
A138_Gt9(92)_Ps		6 Dresden	ST461	Gt 9 (92)	VACC11	Nablus, Palestine	2014	Hopital environmental swab	3,363,143	3,038	162,571	4.8	2,996	40	Alcoy branch
A112_Gt10(93)_Ps		6 Dresden	ST461	Gt 10 (93)	VACC11	Beit Jala, Palestine	2013	Hopital environmental swab	3,362,286	3,030	147,968	4.4	2,990	42	Highly abundant genotype
A114_Gt10(93)_Ps		6 Dresden	ST461	Gt 10 (93)	VACC11	Beit Jala, Palestine	2013	Hospital hot water sample	3,362,313	3,033	159,647	4.7	2,989	41	Highly abundant genotype

A127_Gt10(93)_Ps		6 Dresden	ST461	Gt 10 (93)	VACC11	Beit Jala, Palestine	2013	Hopital environmental swab	3,363,025	3,037	142,978	4.3	2,993	38	Highly abundant genotype
A108_Gt10(93)_Ps		6 Dresden	St461	Gt 10 (93)	VACC11	Beit Jala, Palestine	2013	Hopital environmental swab	3,291,620	2,949	153,342	4.7	2,907	38	Highly abundant genotype
A166_Gt8(142)_Ps		Sg.(2-14)	ST1482	Gt 8 (142)	VACC5	Hebron, Palestine	2014	Hopital environmental swab	3,353,128	3,014	204,232	6.1	2,974	28	Alcoy branch
H34_Gt22(102)_D		Sg.10	NA	Gt 22 (102)		Braunschweig, Germany	2013	Environmental sample	3,403,136	3,049	160,329	4.7	3,016	64	D7630 branch
H29_Gt22(100)_D		4 Portland	NA	Gt 22 (100)	VACC2	Braunschweig, Germany	2013	Environmental sample	3,403,391	3,061	144,415	4.2	3,017	68	D7630 branch
H35_Gt22(102)_D		Sg.1	NA	Gt 22 (102)	VACC5	Braunschweig, Germany	2013	Environmental sample	3,402,267	3,049	166,533	4.9	3,013	64	D7630 branch
H39_Gt4(17)_D		4 Portland	NA	Gt 4 (17)	NA	Braunschweig, Germany	2014	Environmental sample	3,592,274	3,245	226,268	6.3	3,202	78	Worldwide present and highest abundance in our collections
H3_Gt14(31)_D		6 Chicago	ST1431	Gt 14 (31)	VACC2	Braunschweig, Germany	2009	Environmental sample	3,691,263	3,386	209,867	5.7	3,336	2	Alcoy branch
H23_Gt14(30)_D		6 Chicago	ST1431	Gt 14 (30)	VACC2	Braunschweig, Germany	2013	Environmental sample	3,783,853	3,482	372,750	9.9	3,444	130	Alcoy branch
H1_Gt14(31)_D		6 Chicago	ST1431	Gt 14 (31)	VACC2	Braunschweig, Germany	2009	Environmental sample	3,654,038	3,355	253,073	6.9	3,312	69	Alcoy branch

H2_Gt14(31)_D		6 Chicago	ST1431	Gt 14 (31)	VACC2	Braunschweig, Germany	2009	Environmental sample	3,650,216	3,347	238,649	6.5	3,310	64	Alcoy branch
L09-313_Cl_Gt84(116)_D		Sg.3	ST93	Gt 84 (116)	VACC13	Freiburg, Germany	2009	Clinical sample	3,398,240	3,068	160,181	4.7	3,027	58	Thunderbay branch
L01-443_Cl_Gt64(74)_D		1 Knoxville	ST9	Gt 64 (74)	VACC13	Herford, Germany	2001	Clinical sample	3,477,581	3,142	225,639	6.5	3,099	103	Thunderbay branch
L10-091_Cl_Gt69(69)_D		1 Philadelphia	ST435	Gt 69 (69)	VACC2	Erlangen, Germany	2010	Clinical sample	3,687,459	3,373	187,440	5.1	3,330	97	Pontiac Branch
L12-317_Cl_Gt72(66)_D		1 Knoxville	ST444	Gt 72 (66)	VACC2	Lorrach, Germany	2012	Clinical sample	3,516,737	3,170	186,733	5.3	3,133	74	Pontiac Branch
L11-209_Cl_Gt29(27)_D		1 Knoxville	ST62	Gt 29 (27)	VACC2	Hannover, Germany	2011	Clinical sample	3,461,996	3,130	96,408	2.8	3,085	65	Pontiac Branch
L10-023_Cl_Gt75(49)_D		1 Knoxville	ST62	Gt 75 (49)	VACC2	Ulm, Germany	2010	Clinical sample	3,641,882	3,427	NA	NA	3,377	1	Pontiac Branch
L09-329_Cl_Gt75(49)_D		1 Philadelphia	ST62	Gt 75 (49)	VACC2	Trier, Germany	2009	Clinical sample	3,445,544	3,108	117,416	3.4	3,068	76	Pontiac Branch
L02-521_Cl_Gt29(27)_D		1 Philadelphia	ST62	Gt 29 (27)	VACC2	Bad Langensalza, Germany	2002	Clinical sample	3,478,585	3,161	118,405	3.4	3,117	72	Pontiac Branch
L02-465_Cl_Gt27(133)_D		1 Benidorm	ST425	Gt 27(133)	VACC1	Berlin, Germany	2002	Clinical sample	3,500,734	3,173	214,498	6.1	3,135	52	Alcoy branch
L04-545_Cl_Gt40(47)_D		6 Dresden	ST292	Gt 40 (47)	VACC2	Heide-West, Germany	2004	Clinical sample	3,358,959	3,006	160,642	4.8	2,964	34	Alcoy branch, close to Warstein outbreak

L06-153_Cl_Gt71 (135)_D		1 OLDA	ST169	Gt 71 (135)	VACC18	Brandenburg , Germany	2006	Clinical sample	3,405,980	3,074	NA	NA	3,030	39	Alcoy branch
L06-129_Cl_Gt71 (135)_D		1 OLDA	ST169	Gt 71 (135)	VACC18	Brandenburg , Germany	2006	Clinical sample	3,406,112	3,075	172,495	5.1	3,031	40	Alcoy branch
L04-041_Cl_Gt30 (137)_D		Sg.3	ST87	Gt 30 (137)	NA	Kassel, Germany	2004	Clinical sample	3,522,671	3,206	258,174	7.3	3,165	50	Alcoy branch
L05-341_Cl_Gt8(132)_D		6 Chicago	ST81	Gt 8 (132)	VACC16	Darmstadt, Germany	2005	Clinical sample	3,410,856	3,071	219,558	6.4	3,035	31	Alcoy branch
[1] Serogruop (monoclonal antibody)			[2] Sequence Type		[3] Multi-Locus of Variable Number of Tandem Repeats Analysis using 13 loci				[3'] VNTR-Analysis -Clonal Complex				* Coding DNA Sequence		

Table A4: Whole-genome SNP comparison of *L. pneumophila* isolates and reference strains.

L. pneumophila Isolates	Lpn-LPE509	Lpn-Phildelphia1	**Lpn-Thunderbay**	Lpn-ATCC43290	Lpn-lpm7613	Lpn-Lens	**Lpn-Lorraine**	Lpn-HL06041035	**Lpn-D7631**	**Lpn-D7630**	Lpn-D7632	Lpn-Paris	**Lpn-OLDA**	**Lpn-Pontiac**	Lpn-Toronto	Lpn-Alcoy	Lpn-Corby	**A194-Gt40(47)-Ps**	**H3-Gt14(31)-D**
Lpn-LPE509	0	7,893	6,607	4,432	5,681	61,660	52,372	55,729	51,231	51,231	51,232	54,130	54,189	54,867	57,050	56,175	56,022	56,954	56,967
Lpn-Phildelphia1	7,949	0	7,173	6,931	8,114	60,513	52,197	57,012	54,113	54,113	54,114	56,508	56,567	57,130	58,655	59,954	59,707	59,815	59,716
L09-313_Cl_Gt84(116)_D	7,497	6,725	**3,463**	3,736	4,614	60,800	52,184	58,590	52,244	52,244	52,245	59,151	59,210	57,112	58,850	56,022	56,099	56,924	57,071
L01-443_Cl_Gt64(74)_D	9,711	8,315	**6,283**	6,589	7,259	60,203	52,162	58,289	52,019	52,019	52,020	59,314	59,373	57,215	59,036	56,337	56,576	57,346	57,053
Lpn-Thunderbay	6,645	7,122	0	2,376	3,271	60,903	52,499	57,766	50,688	50,688	50,689	58,421	58,480	57,208	58,841	56,525	56,277	57,458	57,151
Lpn-ATCC43290	4,471	6,876	2,376	0	1,663	61,037	52,710	57,620	50,680	50,680	50,681	58,102	58,161	56,886	58,502	56,719	56,690	57,633	57,324
Lpn-lpm7613	5,727	8,062	3,270	1,665	0	60,217	52,715	57,852	51,426	51,426	51,427	58,455	58,518	57,005	58,682	56,692	56,664	57,919	57,438
A156_Gt64(74)_Ps	6,255	5,700	**3,018**	3,288	4,185	61,079	52,328	58,651	50,720	50,720	50,721	59,348	59,407	57,387	58,767	55,986	55,492	57,772	56,644
A129_Gt64(74)_Ps	6,251	5,700	**3,018**	3,288	4,185	61,078	52,328	58,651	50,720	50,720	50,721	59,348	59,407	57,387	58,767	55,986	55,492	57,772	56,644
Lpn-Lens	62,174	60,985	61,462	61,608	60,755	0	55,607	68,139	66,603	66,603	66,604	67,884	67,917	62,288	63,962	70,708	71,182	70,922	70,076
Lpn-Lorraine	52,252	52,094	52,408	52,634	52,624	54,968	0	50,857	52,019	52,019	52,020	46,520	46,547	40,566	44,745	53,380	53,848	53,596	53,236
A15_Gt12(84)_Ps	53,544	52,817	52,966	53,275	53,208	52,506	**13,180**	53,553	54,043	54,043	54,044	50,895	50,922	48,035	49,236	55,931	56,527	56,183	55,605

Lpn-HL06041035	55,334	56,644	57,431	57,284	57,502	66,966	50,330	0	28,898	28,898	28,899	32,907	32,968	39,435	40,855	47,898	48,866	48,578	45,746
Lpn-D7631	50,817	53,705	50,323	50,325	51,070	65,514	51,655	29,003	0	0	1	43,822	43,887	46,214	46,966	45,673	46,299	46,044	44,305
Lpn-D7630	50,817	53,705	50,323	50,325	51,070	65,514	51,655	29,003	0	0	1	43,822	43,887	46,214	46,966	45,673	46,299	46,044	44,305
Lpn-D7632	50,818	53,706	50,324	50,326	51,071	65,515	51,656	29,004	1	1	0	43,823	43,888	46,215	46,967	45,674	46,300	46,045	44,306
H34_Gt22(102)_D	53,856	58,095	54,717	54,848	54,850	66,807	51,795	25,429	**9,793**	**9,793**	9,794	43,034	43,095	44,852	45,631	43,272	43,761	44,291	41,670
H29_Gt22(100)_D	53,855	58,066	54,688	54,819	54,821	66,778	51,766	25,400	**9,764**	**9,764**	9,765	43,005	43,066	44,823	45,602	43,243	43,732	44,262	41,641
H35_Gt22(102)_D	53,855	58,132	54,750	54,885	54,887	66,844	51,832	25,466	**9,831**	**9,831**	9,832	43,071	43,132	44,889	45,669	43,310	43,799	44,329	41,708
H39_Gt4(17)_D	55,366	57,545	57,882	57,756	57,893	62,815	47,911	36,501	44,709	44,709	44,710	11,084	**11,020**	32,760	32,365	46,510	46,595	46,619	46,258
A139_Gt4(17)_Ps	53,876	56,316	58,267	57,951	58,294	66,870	46,172	33,110	43,893	43,893	43,894	260	**195**	28,098	29,498	46,344	46,958	46,574	46,185
Lpn-Paris	53,754	56,173	58,128	57,812	58,155	66,767	46,068	32,935	43,744	43,744	43,745	0	90	28,057	29,457	46,188	46,802	46,417	46,033
A5_Gt4(17)_Ps	53,881	56,289	58,244	57,928	58,271	66,851	46,147	33,054	43,863	43,863	43,864	147	**76**	28,080	29,480	46,305	46,919	46,534	46,150
Lpn-OLDA	53,817	56,232	58,187	57,871	58,214	66,794	46,092	32,997	43,807	43,807	43,808	89	0	28,026	29,424	46,247	46,861	46,476	46,092
A29_Gt6(18)_Ps	53,003	56,351	57,396	57,080	57,411	66,653	46,114	32,784	42,711	42,711	42,712	1,753	**1,688**	28,114	29,416	45,065	45,636	45,379	44,982
A131_Gt6(18)_Ps	53,001	56,347	57,392	57,076	57,407	66,651	46,112	32,782	42,709	42,709	42,710	1,751	**1,686**	28,112	29,414	45,061	45,632	45,375	44,978
L10-091_Cl_Gt69(69)_D	54,313	56,966	57,043	56,697	56,892	64,690	42,654	38,454	44,541	44,541	44,542	24,407	24,377	**14,358**	15,657	45,825	45,796	46,448	46,077
L12-317_Cl_Gt72(66)_D	54,279	56,915	56,951	56,621	56,801	64,191	42,018	38,499	44,954	44,954	44,955	25,422	25,392	**14,949**	16,587	45,264	45,257	46,004	45,589

L11-209_Cl_Gt29(27)_D	54,688	56,911	56,569	56,259	56,280	60,112	40,762	40,236	46,186	46,186	46,187	28,696	28,667	**2,704**	17,498	47,311	47,981	47,912	47,050
L10-023_Cl_Gt75(49)_D	54,483	56,779	56,797	56,483	56,586	61,117	40,207	39,537	46,013	46,013	46,014	27,898	27,869	**823**	16,393	46,774	47,339	47,539	46,453
L09-329_Cl_Gt75(49)_D	54,327	56,492	56,579	56,342	56,445	59,975	40,316	40,251	46,761	46,761	46,762	29,248	29,219	**2,329**	17,797	47,545	47,965	48,201	47,333
Lpn-Pontiac	54,350	56,657	56,744	56,430	56,533	61,016	39,951	39,432	45,999	45,999	46,000	27,956	27,927	0	16,484	46,973	47,501	47,597	46,747
L02-521_Cl_Gt29(27)_D	54,215	56,521	56,608	56,294	56,397	60,051	40,083	40,073	46,572	46,572	46,573	29,082	29,053	**2,030**	17,670	47,787	48,273	48,363	47,595
Lpn-Toronto	56,626	58,242	58,470	58,127	58,294	62,885	44,296	40,967	46,900	46,900	46,901	29,477	29,446	16,653	0	46,576	47,001	47,176	46,853
L02-465_Cl_Gt27(133)_D	56,947	59,635	57,486	57,654	57,718	67,881	51,844	43,966	39,939	39,938	39,939	43,405	43,466	45,846	46,267	19,764	20,122	**19,031**	21,146
L04-545_Cl_Gt40(47)_D	56,226	59,148	56,602	56,796	57,088	69,342	52,781	48,072	46,337	46,336	46,337	45,808	45,869	47,918	46,700	11,776	12,448	**2,822**	15,112
A193_Gt40(47)_Ps	56,790	59,399	57,082	57,269	57,542	69,601	53,182	48,816	46,127	46,126	46,127	46,656	46,717	48,220	47,519	13,008	14,269	**595**	16,883
A194_Gt40(47)_Ps	55,980	58,832	56,516	56,703	56,976	69,038	52,613	48,246	45,547	45,546	45,547	46,086	46,147	47,651	46,952	12,623	13,696	0	16,319
A195_Gt40(47)_Ps	56,044	58,896	56,580	56,767	57,040	69,101	52,677	48,309	45,612	45,611	45,612	46,150	46,211	47,715	47,016	12,686	13,759	**69**	16,383
A138_Gt9(92)_Ps	55,387	58,320	56,257	56,577	56,771	69,093	52,304	47,447	46,595	46,594	46,595	45,200	45,262	47,414	47,230	11,674	12,977	**9,235**	15,497
A112_Gt10(93)_Ps	55,391	58,287	56,228	56,579	56,773	69,035	52,119	47,331	46,562	46,561	46,562	45,023	45,085	47,006	46,832	12,080	13,387	**9,737**	15,668

A114_Gt10(93)_Ps	55,397	58,293	56,234	56,585	56,779	69,041	52,125	47,337	46,568	46,567	46,568	45,029	45,091	47,012	46,838	12,086	13,393	**9,743**	15,674
A127_Gt10(93)_Ps	55,398	58,324	56,265	56,616	56,810	69,073	52,157	47,369	46,600	46,599	46,600	45,061	45,123	47,044	46,870	12,119	13,426	**9,776**	15,707
A108_Gt10(93)_Ps	55,400	58,296	56,237	56,588	56,782	69,044	52,128	47,340	46,571	46,570	46,571	45,032	45,094	47,015	46,841	12,089	13,396	**9,746**	15,677
H3_Gt14(31)_D	55,599	58,312	55,793	55,969	56,073	67,765	51,960	45,184	43,502	43,501	43,502	45,499	45,560	46,694	46,414	12,709	13,932	15,997	0
H23_Gt14(30)_D	55,605	58,318	55,799	55,975	56,079	67,771	51,966	45,190	43,508	43,507	43,508	45,505	45,566	46,700	46,420	12,715	13,938	16,003	**8**
H1_Gt14(31)_D	55,602	58,315	55,796	55,972	56,076	67,768	51,963	45,187	43,505	43,504	43,505	45,502	45,563	46,697	46,417	12,712	13,935	16,000	**5**
H2_Gt14(31)_D	55,602	58,315	55,796	55,972	56,076	67,768	51,963	45,187	43,505	43,504	43,505	45,502	45,563	46,697	46,417	12,712	13,935	16,000	**5**
Lpn-Alcoy	55,191	58,936	55,551	55,754	55,715	68,796	52,373	47,589	45,173	45,172	45,173	45,889	45,950	47,059	46,349	0	8,536	12,575	12,992
Lpn-Corby	54,760	58,420	55,226	55,457	55,428	69,105	52,667	48,358	45,614	45,613	45,614	46,311	46,372	47,418	46,605	8,569	0	13,669	14,240
L06-153_Cl_Gt71(135)_D	56,435	60,030	56,786	57,052	57,085	70,197	53,883	49,343	46,579	46,578	46,579	46,688	46,749	48,436	47,569	11,960	12,203	**11,755**	15,637
L06-129_Cl_Gt71(135)_D	56,432	60,027	56,783	57,049	57,082	70,194	53,880	49,340	46,576	46,575	46,576	46,685	46,746	48,433	47,566	11,957	12,200	**11,752**	15,634
A166_Gt8(142)_Ps	56,734	59,717	57,005	57,273	57,506	69,648	53,331	48,134	44,949	44,948	44,950	45,390	45,451	47,608	46,775	13,650	13,572	**11,522**	16,010
L04-041_Cl_Gt30(137)_D	55,304	58,994	55,498	55,768	55,990	68,718	52,568	45,987	42,328	42,327	42,328	45,102	45,163	46,619	46,345	14,500	14,015	**13,889**	15,584
L05-341_Cl_Gt8(132)_D	56,279	59,240	56,383	56,663	56,758	69,406	52,920	47,086	45,076	45,075	45,076	45,359	45,418	47,058	46,395	13,146	13,809	**11,468**	14,975

A color represents relative number of SNPs differences. Each branch has specific color code as given in Figure 1. **Bold** and Underlined = Least No. of SNPs

Table A5: Summary of Genome Islands of *L.pneumophila* of the selected strains.						
Cluster	Isolate	No of GI	GI	GI length (bp)	No of Genes	Total No. of Genes in GI
ATCC43290 Branch 1	L09-313_Cl_Gt84(116)_D	8	1	6,666	7	162
			2	6,819	10	
			3	16,241	19	
			4	17,924	21	
			5	64,086	64	
			6	16,989	16	
			7	12,159	17	
			8	17,497	8	
	L01-443_Cl_Gt64(74)_D	11	1	25,819	33	214
			2	12,159	17	
			3	6,666	7	
			4	8,619	10	
			5	14,609	16	
			6	11,582	15	
			7	16,992	16	
			8	53,658	60	
			9	13,729	8	
			10	9,075	10	
			11	52,731	22	
	A156_Gt64(74)_Ps	9	1	6,666	7	189
			2	8,619	10	
			3	16,241	19	
			4	16,856	20	
			5	64,062	64	
			6	16,989	16	
			7	21,177	24	
			8	11,562	12	
			9	12,159	17	
	A129_Gt64(74)_Ps	13	1	6,666	7	282
			2	8,619	10	
			3	16,241	19	
			4	16,856	20	
			5	64,062	64	
			6	16,989	16	
			7	18,290	22	
			8	11,562	12	
			9	5,373	8	
			10	48,997	23	
			11	12,225	13	
			12	59,011	58	
			13	9,013	10	

Lorraine Branch 2	A15_Gt12(84)_Ps	11	1	8,007	8	169
			2	13,279	15	
			3	12,162	10	
			4	13,368	18	
			5	11,447	14	
			6	13,771	12	
			7	24,364	24	
			8	7,312	8	
			9	13,264	18	
			10	8,860	11	
			11	54,179	31	
D7630 Branch 3	H34_Gt22(102)_D	8	1	8,009	8	152
			2	16,952	19	
			3	7,152	8	
			4	27,517	24	
			5	32,779	36	
			6	16,619	13	
			7	21,177	24	
			8	30,124	20	
	H29_Gt22(100)_D	8	1	8,009	8	142
			2	16,952	19	
			3	7,152	8	
			4	27,517	24	
			5	37,893	39	
			6	16,619	13	
			7	21,177	24	
			8	9,146	7	
	H35_Gt22(102)_D	10	1	8,009	8	150
			2	16,952	19	
			3	7,152	8	
			4	7,769	9	
			5	25,463	19	
			6	37,843	39	
			7	16,619	13	
			8	21,177	24	
			9	9147	7	
			10	16,402	4	
Paris Branch 4	H39_Gt4(17)_D	11	1	8372	8	174
			2	4699	7	
			3	12,218	10	
			4	11,507	12	
			5	46,869	46	
			6	6656	7	
			7	5238	7	
			8	43,618	37	
			9	59,039	11	
			10	8497	2	
			11	19,555	27	

Paris Branch 4	A139_Gt4(17)_Ps	11	1	8372	8	224
			2	4699	7	
			3	13,220	14	
			4	22,127	23	
			5	5310	7	
			6	43,012	37	
			7	6954	11	
			8	46,869	46	
			9	6656	7	
			10	50,258	54	
			11	16,374	10	
	A5_Gt4(17)_Ps	10	1	4699	7	215
			2	12,039	11	
			3	22,127	23	
			4	5310	7	
			5	43,012	37	
			6	6954	11	
			7	46,869	46	
			8	6656	7	
			9	49,633	53	
			10	28,689	13	
	A29_Gt6(18)_Ps	12	1	8372	8	309
			2	4699	7	
			3	12,064	11	
			4	22,127	23	
			5	27,526	24	
			6	30,804	35	
			7	42,942	37	
			8	6954	11	
			9	74,559	79	
			10	6656	7	
			11	51,378	55	
			12	50,924	12	
	A131_Gt6(18)_Ps	13	1	8372	8	268
			2	4699	7	
			3	12,041	11	
			4	22,127	23	
			5	27,526	24	
			6	30,804	35	
			7	8807	8	
			8	42,942	37	
			9	6954	11	
			10	20,474	17	
			11	64,866	73	
			12	6656	7	
			13	24,951	7	

Pontiac Branch 5	L10-091_Cl_Gt 69(69)_D	13	1	5520	8	175
			2	12,846	11	
			3	8997	7	
			4	6739	11	
			5	26,673	37	
			6	9527	13	
			7	8836	9	
			8	20,964	21	
			9	15,724	24	
			10	8519	11	
			11	8920	7	
			12	6656	7	
			13	47,519	9	
	L12-317_Cl_Gt 72(66)_D	10	1	32,465	28	176
			2	12,224	9	
			3	12,170	10	
			4	64,066	64	
			5	19,685	20	
			6	6014	7	
			7	13,114	15	
			8	11,419	9	
			9	8920	7	
			10	6656	7	
	L11-209_Cl_Gt 29(27)_D	6	1	17,154	18	94
			2	12,846	11	
			3	27,009	29	
			4	15,752	11	
			5	16,990	18	
			6	6657	7	
	L09-329_Cl_Gt 75(49)_D	7	1	6657	7	116
			2	46,585	41	
			3	6867	11	
			4	12,857	11	
			5	8577	7	
			6	5511	7	
			7	30,362	32	
	L02-521_Cl_Gt 29(27)_D	8	1	13,937	17	108
			2	6867	11	
			3	9929	8	
			4	6890	7	
			5	8721	8	
			6	39,879	42	
			7	6657	7	
			8	25,525	8	

Alcoy Branch 6	L02-465_Cl_Gt 27(133)_D	13	1	4180	7	215
			2	9993	8	
			3	9620	8	
			4	5031	7	
			5	27,758	24	
			6	65,316	71	
			7	8431	9	
			8	21,716	29	
			9	5477	7	
			10	14,122	13	
			11	17,732	20	
			12	6667	7	
			13	18,455	5	
	L04-545_Cl_Gt 40(47)_D	8	1	6667	7	144
			2	20,340	30	
			3	16,142	16	
			4	7119	7	
			5	55,223	30	
			6	28,334	25	
			7	9085	9	
			8	17,732	20	
	A193_Gt40(47)_Ps	8	1	6667	7	161
			2	19,317	27	
			3	17,424	17	
			4	7119	7	
			5	68,131	36	
			6	28,258	25	
			7	19,749	19	
			8	18,487	23	
	A194_Gt40(47)_Ps	6	1	15,197	1	108
			2	17,424	16	
			3	68,995	35	
			4	28,295	27	
			5	9085	9	
			6	17,732	20	
	A195_Gt40(47)_Ps	8	1	6667	7	146
			2	19,317	27	
			3	16,142	15	
			4	7119	7	
			5	26,090	33	
			6	28,334	25	
			7	9085	9	
			8	18,487	23	
	A138_Gt9(92)_Ps	7	1	6667	7	127
			2	37,408	46	
			3	59,978	34	
			4	8980	7	
			5	8429	8	
			6	17,721	20	
			7	23,388	5	

Alcoy Branch 6	A112_Gt10(93)_Ps	6	1	17,721	20	120
			2	6667	7	
			3	37,408	46	
			4	58,777	32	
			5	8429	8	
			6	18,966	7	
	A114_Gt10(93)_Ps	6	1	6667	7	122
			2	37,408	46	
			3	58,777	32	
			4	8429	8	
			5	17,721	20	
			6	30,645	9	
	A127_Gt10(93)_Ps	6	1	6667	7	127
			2	42,945	56	
			3	58,777	32	
			4	8429	8	
			5	17,721	20	
			6	8439	4	
	A108_Gt10(93)_Ps	7	1	6667	7	134
			2	42,991	55	
			3	58,782	32	
			4	7941	8	
			5	8429	8	
			6	17,721	20	
			7	10,811	4	
	H1_Gt14(31)_D	10	1	6668	7	226
			2	13,062	16	
			3	5996	10	
			4	15,575	17	
			5	60,003	35	
			6	64,648	67	
			7	32,103	31	
			8	8437	8	
			9	19,754	21	
			10	26,827	14	
	H2_Gt14(31)_D	11	1	13,062	16	211
			2	5996	10	
			3	15,575	17	
			4	60,053	35	
			5	45,783	49	
			6	9077	11	
			7	32,103	31	
			8	8437	8	
			9	19,811	21	
			10	6668	7	
			11	22,084	6	

Alcoy Branch 6	H3_Gt14(31)_D	12	1	23,425	25	198
			2	13,062	16	
			3	5996	10	
			4	12,559	17	
			5	6615	11	
			6	65,118	36	
			7	10,222	9	
			8	27,708	30	
			9	9978	13	
			10	10,560	8	
			11	14,861	13	
			12	9763	10	
	H23_Gt14(30)_D	13	1	13,062	16	315
			2	5996	10	
			3	15,575	17	
			4	5964	7	
			5	21,994	34	
			6	64,640	36	
			7	45,836	47	
			8	12,096	13	
			9	40,813	43	
			10	8437	8	
			11	19,772	21	
			12	6668	7	
			13	111,897	56	
	L06-153_Cl_Gt71(135)_D	NA	NA	NA	NA	NA
	L06-129_Cl_Gt71(135)_D	5	1	63,906	65	167
			2	54,896	59	
			3	18,290	22	
			4	27,397	11	
			5	8006	10	
	A166_Gt8(142)_Ps	11	1	6667	7	188
			2	13,958	17	
			3	14,858	14	
			4	9620	8	
			5	15,564	12	
			6	21,177	24	
			7	60,115	33	
			8	21,716	29	
			9	12,520	13	
			10	10,305	11	
			11	17,732	20	

Alcoy Branch 6	L04-041_Cl_Gt 30(137)_D	14	1	6667	7	245
			2	10,111	9	
			3	9620	8	
			4	27,517	24	
			5	31,931	36	
			6	7390	10	
			7	7228	10	
			8	21,177	24	
			9	14,752	12	
			10	61,045	34	
			11	15,724	24	
			12	12,529	13	
			13	14,751	14	
			14	17,732	20	
	L05-341_Cl_Gt 8(132)_D	11	1	6667	7	205
			2	24,616	22	
			3	11,536	9	
			4	9620	8	
			5	18,192	24	
			6	21,177	24	
			7	60,169	33	
			8	21,716	29	
			9	12,532	13	
			10	15,601	16	
			11	17,732	20	
NA: Not Available						

Table A6 (Part 1): Percentage of nucleotide identity of orthologous pore-forming activity genes with respect to the BLASTp search against the VFDB using *L. pneumophila* strain Philadelphia1 as default reference genome

***L. pneumophila* isolate**	***Secretion system***											
	Dot/Icm type IVB secretion system											
	icmT			***icmS***			***icmR***			***icmQ***		
	Locus	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**
Lpn-Corby	LPC_2902	82%	260	LPC_2901	100%	344	LPC_2900	95%	362	LPC_2899	100%	575
Lpn-Alcoy 2300/99	Alcoy_00488	82%	260	Alcoy_0489	100%	344	Alcoy_00490	95%	362	Alcoy_00491	100%	575
L02-465_Cl_Gt27(133)_D	L02_465-D-cl_01280	82%	260	L02_465-D-cl_01281	100%	344	L02_465-D-cl_01282	95%	362	L02_465-D-cl_01284	100%	575
L04-545_Cl_Gt40(47)_D	L04_545_00300	82%	260	L04_545_00301	100%	344	L04_545_00302	95%	362	L04_545_00303	100%	575
A193_Gt40(47)_Ps	A193_00315	82%	260	A193_00314	100%	344	A193_00313	95%	362	A193_00312	100%	575
A194_Gt40(47)_Ps	A194_00442	82%	260	A194_00443	100%	344	A194_00444	95%	362	A194_00445	100%	575
A195_Gt40(47)_Ps	A195_01239	82%	260	A195_01238	100%	344	A195_01237	95%	362	A195_01236	100%	575
A138_Gt9(92)_Ps	A138_01248	82%	260	A138_01247	100%	344	A138_01246	95%	362	A138_01245	100%	575
A112_Gt10(93)_Ps	A112_01239	82%	260	A112_01240	100%	344	A112_01241	95%	362	A112_01242	100%	575
A114_Gt10(93)_Ps	A114_01459	82%	260	A114_01458	100%	344	A114_01457	95%	362	A114_01456	100%	575
A127_Gt10(93)_Ps	A127_00229	82%	260	A127_00230	100%	344	A127_00231	95%	362	A127_00232	100%	575

A108_Gt10(93)_Ps	A108_00245	82%	260	A108_00246	100%	344	A108_00247	95%	362	A108_00248	100%	575
H3_Gt14(31)_D	H3_00490	82%	260	H3_00491	100%	344	H3_00492	95%	362	H3_00493	100%	575
H23_Gt14(30)_D	H23_02676	82%	260	H23_02677	100%	344	H23_02678	95%	362	H23_02679	100%	575
H1_Gt14(31)_D	H1_01840	82%	260	H1_01841	100%	344	H1_01842	95%	362	H1_01843	100%	575
H2_Gt14(31)_D	H2_03184	82%	260	H2_03185	100%	344	H2_03186	95%	362	H2_03187	100%	575
L06-153_Cl_Gt71(135)_D	L06_153_02963	82%	260	L06_153_02962	100%	344	L06_153_02961	95%	362	L06_153_02960	100%	575
L06-129_Cl_Gt71(135)_D	L06_129_01414	82%	260	L06_129_01413	100%	344	L06_129_01412	95%	362	L06_129_01411	100%	575
A166_Gt8(142)_Ps	A166_00456	82%	260	A166_00457	100%	344	A166_00458	95%	362	A166_00459	100%	575
L04-041_Cl_Gt30(137)_D	L04_041_03094	82%	260	L04_041_03093	100%	344	L04_041_03092	95%	362	L04_041_03091	100%	575
L05-341_Cl_Gt8(132)_D	L05_341_02188	82%	260	L05_341_02189	100%	344	L05_341_02190	95%	362	L05_341_02191	100%	575
Lpn-Pontiac	Pontiac_00497	82%	260	Pontiac_00498	100%	344	Pontiac_00499	96%	362	Pontiac_00500	99%	575
L10-091_Cl_Gt69(69)_D	L10_091_01392	82%	260	L10_091_01393	100%	344	L10_091_01394	96%	362	L10_091_01395	99%	575
L12-317_Cl_Gt72(66)_D	L12_317_02108	82%	260	L12_317_02107	100%	344	L12_317_02106	96%	362	L12_317_02105	99%	575

L11-209_Cl_Gt29(27)_D	L11_209_00706	82%	260	L11_209_00705	100%	344	L11_209_00704	96%	362	L11_209_00703	99%	575
L10-023_Cl_Gt75(49)_D	L10_023_00511	82%	260	L10_023_00512	100%	344	L10_023_00513	96%	362	L10_023_00514	99%	575
L09-329_Cl_Gt75(49)_D	L09_329_00128	82%	260	L09_329_00127	100%	344	L09_329_00126	96%	362	L09_329_00125	99%	575
L02-521_Cl_Gt29(27)_D	L02_521_00369	82%	260	L02_521_00368	100%	344	L02_521_00367	96%	362	L02_521_00366	99%	575
Lpn-Paris	Paris_00513	82%	260	Paris_00514	100%	344	Paris_00515	96%	362	Paris_00516	99%	575
H39_Gt4(17)_D	H39_00241	82%	260	H39_00242	100%	344	H39_00243	96%	362	H39_00244	99%	575
A139_Gt4(17)_Ps	A139_01748	82%	260	A139_01749	100%	344	A139_01750	96%	362	A139_01751	99%	575
A5_Gt4(17)_Ps	A5_01374	82%	260	A5_01373	100%	344	A5_01372	96%	362	A5_01371	99%	575
A29_Gt6(18)_Ps	A29_01469	82%	260	A29_01470	100%	344	A29_01471	96%	362	A29_01472	99%	575
A131_Gt6(18)_Ps	A131_01819	82%	260	A131_01818	100%	344	A131_01817	96%	362	A131_01816	99%	575
Lpn-D7630	Lpn_D7631_00446	82%	260	lpn_D7631_00447	99%	344	Lpn_D7630_00448	97%	362	lpn_D7630_00449	99%	575
H34_Gt22(102)_D	H34_01933	82%	260	H34_01934	99%	344	H34_01935	97%	362	H34_01936	99%	575
H29_Gt22(100)_D	H29_00805	82%	260	H29_00804	99%	344	H29_00803	97%	362	H29_00802	99%	575
H35_Gt22(102)_D	H35_00748	82%	260	H35_00747	99%	344	H35_00746	97%	362	H35_00745	99%	575

Lpn-Lorraine	Lorriane_00478	82%	260	Lorriane_00479	100%	344	Lorriane_00480	97%	362	Lorriane_00481	100%	575
A15_Gt12(84)_Ps	A15_00834	82%	260	A15_00833	100%	344	A15_00832	97%	362	A15_00831	100%	575
Lpn-Thunderbay	Thunderbay_00431	82%	260	Thunderbay_00432	100%	344	Thunderbay_00433	91%	362	Thunderbay_00434	100%	575
L09-313_Cl_Gt84(116)_D	L09_313_01086	82%	260	L09_313_01087	100%	344	L09_313_01088	91%	362	L09_313_01089	100%	575
L01-443_Cl_Gt64(74)_D	L01_443_01651	82%	260	L01_443_01650	100%	344	L01_443_01649	91%	362	L01_443_01648	100%	575
A156_Gt64(74)_Ps	A156_01687	82%	260	A156_01688	100%	344	A156_01689	91%	362	A156_01690	100%	575
A129_Gt64(74)_Ps	A129_01377	82%	260	A129_01378	100%	344	A129_01379	91%	362	A129_01380	100%	575

Table A6 (Part 2): Percentage of nucleotide identity of orthologous pore-forming activity genes with respect to the BLASTp search against the VFDB using *L. pneumophila* strain Philadelphia1 as default reference genome

***L. pneumophila* isolate**	***Secretion system***											
	Dot/Icm type IVB secretion system											
	icmL/dotI			***icmK/dotH***			***icmE/dotG***			***icmC/dotE***		
	Locus	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Id**	**Id**	**Length (bp)**
Lpn-Corby	LPC_2894	84%	638	LPC_2893	84%	1,082	LPC_2892	88.00%	3,146	LPC_2890	91.00%	341
Lpn-Alcoy 2300/99	Alcoy_00496	84%	638	Alcoy_00497	84%	1,085	Alcoy_00498	88.00%	3,146	Alcoy_00500	99.00%	584
L02-465_Cl_Gt27(133)_D	L02_465-D-cl_01288	84%	638	L02_465-D-cl_01289	84%	1,085	L02_465-D-cl_01290	88.00%	3,146	L02_465-D-cl_01292	99.00%	584
L04-545_Cl_Gt40(47)_D	L04_545_00308	84%	638	L04_545_00309	84%	1,085	L04_545_00310	88.00%	3,146	L04_545_00312	99.00%	584
A193_Gt40(47)_Ps	A193_00307	84%	638	A193_00306	84%	1,085	A193_00305	88.00%	3,146	A193_00302	99.00%	584
A194_Gt40(47)_Ps	A194_00450	84%	638	A194_00451	84%	1,085	A194_00452	88.00%	3,146	A194_00454	99.00%	584
A195_Gt40(47)_Ps	A195_01231	84%	638	A195_01230	84%	1,085	A195_01229	88.00%	3,146	A195_01227	99.00%	584
A138_Gt9(92)_Ps	A138_01240	84%	638	A138_01239	83%	1,085	A138_01238	88.00%	3,146	A138_01236	99.00%	548
A112_Gt10(93)_Ps	A112_01247	84%	638	A112_01248	83%	1,085	A112_01249	88.00%	3,146	A112_01251	99.00%	548
A114_Gt10(93)_Ps	A114_01451	84%	638	A114_01450	83%	1,085	A114_01449	88.00%	3,146	A114_01447	99.00%	548
A127_Gt10(93)_Ps	A127_00237	84%	638	A127_00238	83%	1,085	A127_00239	88.00%	3,146	A127_00241	99.00%	548

A108_Gt10(93)_Ps	A108_00253	84%	638	A108_00254	83%	1,085	A108_00255	88.00%	3,146	A108_00257	99.00%	548
H3_Gt14(31)_D	H3_00498	84%	638	H3_00499	83%	1,085	H3_00500	87.00%	3,146	H3_00502	99.00%	548
H23_Gt14(30)_D	H23_02684	84%	638	H23_02685	83%	1,085	H23_02686	87.00%	3,146	H23_02688	99.00%	584
H1_Gt14(31)_D	H1_01848	84%	638	H1_01849	83%	1,085	H1_01850	87.00%	3,146	H1_01852	99.00%	584
H2_Gt14(31)_D	H2_03192	84%	638	H2_03193	83%	1,085	H2_03194	87.00%	3,146	H2_03196	99.00%	584
L06-153_Cl_Gt71(135)_D	L06_153_02955	84%	638	L06_153_02954	84%	1,085	L06_153_02953	88.00%	3,146	L06_153_02951	99.00%	584
L06-129_Cl_Gt71(135)_D	L06_129_01406	84%	638	L06_129_01405	84%	1,085	L06_129_01404	88.00%	3,146	L06_129_01402	99.00%	584
A166_Gt8(142)_Ps	A166_00464	84%	638	A166_00465	84%	1,085	A166_00466	89.00%	3,146	A166_00469	99.00%	584
L04-041_Cl_Gt30(137)_D	L04_041_03086	84%	638	L04_041_03085	84%	1,085	L04_041_03084	88.00%	3,146	L04_041_03082	99.00%	584
L05-341_Cl_Gt8(132)_D	L05_341_02196	84%	638	L05_341_02197	84%	1,085	L05_341_02198	88.00%	3,146	L05_341_02200	99.00%	584
Lpn-Pontiac	Pontiac_00505	84%	638	Pontiac_00506	82%	1,085	Pontiac_00507	88.00%	3,146	Pontiac_00509	100.00%	584
L10-091_Cl_Gt69(69)_D	L10_091_01400	84%	638	L10_091_01401	82%	1,085	L10_091_01402	88.00%	3,146	L10_091_01404	100.00%	584
L12-317_Cl_Gt72(66)_D	L12_317_02100	84%	638	L12_317_02099	82%	1,085	L12_317_02098	88.00%	3,146	L12_317_02096	100.00%	584

L11-209_Cl_Gt29(27)_D	L11_209_00698	84%	638	L11_209_00697	82%	1,085	L11_209_00696	88.00%	3,146	L11_209_00694	100.00%	584
L10-023_Cl_Gt75(49)_D	L10_023_00519	84%	638	L10_023_00520	82%	1,085	L10_023_00521	88.00%	3,146	L10_023_00523	100.00%	584
L09-329_Cl_Gt75(49)_D	L09_329_00120	84%	638	L09_329_00119	82%	1,085	L09_329_00118	88.00%	3,146	L09_329_00116	100.00%	584
L02-521_Cl_Gt29(27)_D	L02_521_00361	84%	638	L02_521_00360	82%	1,085	L02_521_00359	88.00%	3,146	L02_521_00357	100.00%	584
Lpn-Paris	Paris_00521	84%	638	Paris_00522	82%	1,085	Paris_00523	89.00%	3,146	Paris_00525	100.00%	584
H39_Gt4(17)_D	H39_00249	84%	638	H39_00250	82%	1,085	H39_00251	89.00%	3,146	H39_00253	100.00%	584
A139_Gt4(17)_Ps	A139_01756	84%	638	A139_01757	82%	1,082	A139_01758	89.00%	3,146	A139_01760	100.00%	584
A5_Gt4(17)_Ps	A5_01366	84%	638	A5_01365	82%	1,082	A5_01364	89.00%	3,146	A5_01362	100.00%	584
A29_Gt6(18)_Ps	A29_01477	84%	638	A29_01478	82%	1,082	A29_01479	89.00%	3,146	A29_01481	100.00%	584
A131_Gt6(18)_Ps	A131_01811	84%	638	A131_01810	82%	1,082	A131_01809	89.00%	3,146	A131_01807	100.00%	584
Lpn-D7630	lpn_D7630_00454	84%	638	lpn_D7630_00455	82%	1,085	lpn_D7630_00456	87.00%	3,146	lpn_D7630_00458	99.00%	584
H34_Gt22(102)_D	H34_01941	84%	638	H34_01942	82%	1,085	H34_01943	87.00%	3,146	H34_01945	99.00%	584
H29_Gt22(100)_D	H29_00797	84%	638	H29_00796	82%	1,085	H29_00795	87.00%	3,146	H29_00793	99.00%	584
H35_Gt22(102)_D	H35_00740	84%	638	H35_00739	82%	1,085	H35_00738	87.00%	3,146	H35_00736	99.00%	584

Lpn-Lorraine	Lorriane_00486	84%	638	Lorriane_00487	82%	1,085	Lorriane_00488	87.00%	3,146	Lorriane_00490	100.00%	584
A15_Gt12(84)_Ps	A15_00826	84%	638	A15_00825	82%	1,082	A15_00824	87.00%	3,146	A15_00822	100.00%	584
Lpn-Thunderbay	Thunderbay_00439	84%	638	Thunderbay_00440	79%	1,085	Thunderbay_00441	91.00%	3,146	Thunderbay_00443	100.00%	584
L09-313_Cl_Gt84(116)_D	L09_313_01094	84%	638	L09_313_01095	79%	1,085	L09_313_01096	91.00%	3,146	L09_313_01098	100.00%	584
L01-443_Cl_Gt64(74)_D	L01_443_01643	84%	638	L01_443_01642	79%	1,085	L01_443_01641	91.00%	3,146	L01_443_01639	100.00%	584
A156_Gt64(74)_Ps	A156_01695	84%	638	A156_01696	79%	1,085	A156_01697	91.00%	3,146	A156_01699	100.00%	584
A129_Gt64(74)_Ps	A129_01385	84%	638	A129_01386	79%	1,085	A129_01387	91.00%	3,146	A129_01389	100.00%	584

Table A6 (Part 3): Percentage of nucleotide identity of orthologous pore-forming activity genes with respect to the BLASTp search against the VFDB using *L. pneumophila* strain Philadelphia1 as default reference genome

***L. pneumophila* isolate**	***Secretion system***									***Toxin***		
	Dot/Icm type IVB secretion system									**RtxA**		
	dotB			***dotA***			***icmW***			***rtxA***		
	Locus	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**	**Locus**	**Id**	**Length (bp)**
Lpn-Corby	LPC_0461	100.00%	1,109	LPC_0450	85.00%	3,062	LPC_0448	98.00%	455	LPC_2649	85.00%	18,869
Lpn-Alcoy 2300/99	Alcoy_02816	100.00%	1,133	Alcoy_02826	81.00%	3,059	Alcoy_02828	98.00%	455	Alcoy_00732	85.00%	14,009
L02-465_Cl_Gt27(133)_D	L02_465-D-cl_02396	100.00%	1,133	L02_465-D-cl_02386	81.00%	3,107	L02_465-D-cl_02384	99.00%	455	L02_465_00601	85.00%	3,854
L04-545_Cl_Gt40(47)_D	L04_545_02079	100.00%	1,133	L04_545_02089	81.00%	3,119	L04_545_02091	98.00%	455	L04_545_01005	84.00%	3,827
A193_Gt40(47)_Ps	A193_02294	100.00%	1,133	A193_02304	78.00%	3,119	A193_02306	98.00%	455	A193_02596	84.00%	3,827
A194_Gt40(47)_Ps	A194_02775	100.00%	1,133	A194_02785	78.00%	3,119	A194_02787	98.00%	455	A194_00644	84.00%	15,197
A195_Gt40(47)_Ps	A195_03002	100.00%	1,133	A195_03012	78.00%	3,119	A195_03014	98.00%	455	A195_01355	84.00%	4,331
A138_Gt9(92)_Ps	A138_02711	100.00%	1,133	A138_02701	78.00%	3,119	A138_02699	98.00%	455	A138_00037	75.00%	4,877
A112_Gt10(93)_Ps	A112_01960	100.00%	1,133	A112_01970	78.00%	3,119	A112_01972	98.00%	455	A112_01365	75.00%	4,877
A114_Gt10(93)_Ps	A114_02800	100.00%	1,133	A114_02810	78.00%	3,119	A114_02812	98.00%	455	A114_01666	75.00%	4,877
A127_Gt10(93)_Ps	A127_02015	100.00%	1,133	A127_02005	78.00%	3,119	A127_02003	98.00%	455	A127_00073	75.00%	5,360

A108_Gt10(93)_Ps	A108_02616	100.00%	1,133	A108_02626	78.00%	3,119	A108_02628	98.00%	455	A108_02733	75.00%	4,877
H3_Gt14(31)_D	H3_02919	100.00%	1,133	H3_02929	79.00%	3,119	H3_02931	98.00%	455	H3_00731	86.00%	14,000
H23_Gt14(30)_D	H23_00073	100.00%	1,133	H23_00063	78.00%	3,119	H23_00061	98.00%	455	H23_01993	86.00%	4,892
H1_Gt14(31)_D	H1_01583	100.00%	1,133	H1_01573	78.00%	3,119	H1_01571	98.00%	455	H1_02924	86.00%	4,892
H2_Gt14(31)_D	H2_01436	100.00%	1,133	H2_01446	78.00%	3,119	H2_01448	98.00%	455	H2_03057	86.00%	4,892
L06-153_Cl_Gt71(135)_D	L06_153_02441	100.00%	1,133	L06_153_02451	80.00%	3,059	L06_153_02453	98.00%	455	L06_153_00001	85.00%	3,854
L06-129_Cl_Gt71(135)_D	L06_129_02607	100.00%	1,133	L06_129_02617	80.00%	3,059	L06_129_02619	98.00%	455	L06_129_00121	85.00%	3,854
A166_Gt8(142)_Ps	A166_02937	100.00%	1,133	A166_02947	80.00%	3,059	A166_02949	98.00%	455	A166_01463	84.00%	3,827
L04-041_Cl_Gt30(137)_D	L04_041_01669	100.00%	1,133	L04_041_01679	80.00%	3,059	L04_041_01681	98.00%	455	L04_041_01996	85.00%	3,875
L05-341_Cl_Gt8(132)_D	L05_341_02472	100.00%	1,133	L05_341_02462	80.00%	3,059	L05_341_02460	98.00%	455	L05_341_02062	84.00%	3,881
Lpn-Pontiac	Pontiac_02875	100.00%	1,133	Pontiac_02886	88.00%	890	Pontiac_02888	99.00%	455	Pontiac_00698	80.00%	24,137
L10-091_Cl_Gt69(69)_D	L10_091_01582	100.00%	1,133	L10_091_01593	81.00%	3,107	L10_091_01595	99.00%	455	L10_091_00688	75.00%	4,421
L12-317_Cl_Gt72(66)_D	L12_317_00927	100.00%	1,133	L12_317_00938	81.00%	3,107	L12_317_00940	99.00%	455	L12_317_01355	75.00%	4,421

L11-209_Cl_Gt29(27)_D	L11_209_03041	100.00%	1,133	L11_209_03031	81.00%	3,107	L11_209_03029	99.00%	455	L11_209_01543	80.00%	4,508
L10-023_Cl_Gt75(49)_D	L10_023_03069	100.00%	1,133	L10_023_03079	81.00%	3,107	L10_023_03081	99.00%	455	L10_023_00715	80.00%	7,496
L09-329_Cl_Gt75(49)_D	L09_329_02691	100.00%	1,133	L09_329_02681	81.00%	3,107	L09_329_02679	99.00%	455	L09_329_01602	80.00%	4,508
L02-521_Cl_Gt29(27)_D	L02_521_02549	100.00%	1,133	L02_521_02559	81.00%	3,107	L02_521_02561	99.00%	455	L02_521_01083	80.00%	4,508
Lpn-Paris	Paris_02793	99.00%	1,133	Paris_02803	81.00%	3,107	Paris_02805	99.00%	455	Paris_00715	83.00%	23,039
H39_Gt4(17)_D	H39_03159	100.00%	1,133	H39_03169	81.00%	3,107	H39_03171	99.00%	455	H39_01209	83.00%	5,012
A139_Gt4(17)_Ps	A139_02971	99.00%	1,133	A139_02981	81.00%	3,107	A139_02983	99.00%	455	A139_01118	83.00%	5,012
A5_Gt4(17)_Ps	A5_02597	99.00%	1,133	A5_02587	81.00%	3,107	A5_02585	99.00%	455	A5_03010	83.00%	5,012
A29_Gt6(18)_Ps	A29_02967	99.00%	1,133	A29_02957	81.00%	3,107	A29_02955	99.00%	455	A29_01942	83.00%	5,012
A131_Gt6(18)_Ps	A131_03134	99.00%	1,133	A131_03144	81.00%	3,107	A131_03146	99.00%	455	A131_01368	83.00%	5,012
Lpn-D7630	lpn_D7630_02761	100.00%	1,133	lpn_D7630_02772	81.00%	3,104	lpn_D7630_02774	99.00%	455	lpn_D7630_00651	85.00%	22,649
H34_Gt22(102)_D	H34_02347	100.00%	1,133	H34_00983	81.00%	3,104	H34_00981	99.00%	455	H34_00221	85.00%	4,415
H29_Gt22(100)_D	H29_02329	100.00%	1,133	H29_02016	81.00%	3,104	H29_02018	99.00%	455	H29_00449	85.00%	4,415
H35_Gt22(102)_D	H35_02906	100.00%	1,133	H35_02425	81.00%	3,104	H35_02423	99.00%	455	H35_01986	85.00%	4,415

Lpn-Lorraine	Lorriane_02816	100.00%	1,133	Lorriane_02827	81.00%	3,095	Lorriane_02829	99.00%	455	Lorriane_00680	85.00%	12,950
A15_Gt12(84)_Ps	A15_01027	100.00%	1,133	A15_01015	81.00%	3,095	A15_01013	99.00%	455	A15_02665	77.00%	4,475
Lpn-Thunderbay	Thunderbay_02811	100.00%	1,133	Thunderbay_02821	93.00%	3,146	Thunderbay_02823	100.00%	455	Thunderbay_00633	85.00%	6,989
L09-313_Cl_Gt84(116)_D	L09_313_01999	100.00%	1,133	L09_313_01989	93.00%	3,146	L09_313_01987	100.00%	455	L09_313_00347	85.00%	3,854
L01-443_Cl_Gt64(74)_D	L01_443_01451	100.00%	1,133	L01_443_01461	93.00%	3,146	L01_443_01463	100.00%	455	L01_443_01090	85.00%	4,850
A156_Gt64(74)_Ps	A156_02117	100.00%	1,133	A156_02107	93.00%	3,146	A156_02105	100.00%	455	A156_02430	85.00%	3,875
A129_Gt64(74)_Ps	A129_01186	100.00%	1,133	A129_01176	93.00%	3,146	A129_01174	100.00%	455	A129_03205	85.00%	3,875

Table A7: *L. pneumophila* branch one isolates from the West Bank genomic islands and their gene products					
GI Length (bp)	**Gene ID**	**Locus**	**Gene Length (bp)**	**Product (A129_Gt64(74)_Ps)**	**Product (A156_Gt64(74)_Ps)**
6,666	oxaA	A129_00349	1,670	Preprotein translocase subunit YidC	Preprotein translocase subunit YidC
6,666		A129_00350	227	Putative membrane protein insertion efficiency factor	Putative membrane protein insertion efficiency factor
6,666	rnpA	A129_00351	344	ribonuclease P protein component	ribonuclease P protein component
6,666	rpmH	A129_00352	134	50S ribosomal protein L34	50S ribosomal protein L34
6,666	dnaA	A129_00353	1,358	chromosomal replication initiator protein DnaA	chromosomal replication initiator protein DnaA
6,666	dnaN	A129_00354	1,103	DNA polymerase III beta chain	DNA polymerase III beta chain
6,666	recF	A129_00355	1,061	DNA recombination and repair protein ATPase RecF	DNA recombination and repair protein ATPase RecF
8,619		A129_00411	908	transcriptional regulator, LysR family	transcriptional regulator, LysR family
8,619		A129_00412	719	hypothetical protein	hypothetical protein
8,619		A129_00413	404	putative regulator, MerR family transcription regulator	putative regulator, MerR family transcription regulator
8,619		A129_00414	743	hypothetical protein	hypothetical protein
8,619		A129_00415	1,091	hypothetical protein	hypothetical protein
8,619		A129_00416	281	phage related integrase	phage related integrase
8,619		A129_00417	182	hypothetical protein	hypothetical protein
8,619		A129_00418	665	methylase	methylase
8,619		A129_00419	1,178	multidrug resistance protein	multidrug resistance protein
8,619		A129_00420	1,271	hypothetical protein	hypothetical protein
16,241		A129_00500	1,304	hypothetical protein	hypothetical protein
16,241		A129_00501	659	phage repressor	phage repressor
16,241	lvrA2	A129_00502	884	Legionella vir region protein	Legionella vir region protein
16,241	csrA_2	A129_00503	266	global regulator (carbon storage regulator)	global regulator (carbon storage regulator)

16,241		A129_00504	965	LvhB11	LvhB11
16,241	trbC	A129_00505	377	Conjugal transfer protein trbC	Conjugal transfer protein trbC
16,241	trbD	A129_00506	299	Conjugal transfer protein trbD	Conjugal transfer protein trbD
16,241	trbE	A129_00507	2,537	Conjugal transfer protein trbE precursor	Conjugal transfer protein trbE precursor
16,241	trbF	A129_00508	743	Conjugal transfer protein trbF	Conjugal transfer protein trbF
16,241	trbG	A129_00509	875	conjugal transfer protein trbG precursor	conjugal transfer protein trbG precursor
16,241		A129_00510	410	Conjugal transfer protein TrbH	Conjugal transfer protein TrbH
16,241	trbI	A129_00511	1,241	Conjugal transfer protein trbI	Conjugal transfer protein trbI
16,241	trbJ	A129_00512	740	Conjugal transfer protein trbJ precursor	Conjugal transfer protein trbJ precursor
16,241		A129_00513	212	hypothetical protein	hypothetical protein
16,241	trbL	A129_00514	1472	putative conjugal transfer protein trbL	putative conjugal transfer protein trbL
16,241	traG	A129_00515	1,889	Conjugal transfer protein traG	Conjugal transfer protein traG
16,241	traF	A129_00516	545	Conjugal transfer protein traF precursor	Conjugal transfer protein traF precursor
16,241	traD	A129_00517	164	Protein traD	Protein traD
16,241	traC	A129_00518	2,186	DNA primase traC (Replication primase)	DNA primase traC (Replication primase)
16,856		A129_00527	482	cytidine/deoxycytidylate deaminase	cytidine/deoxycytidylate deaminase
16,856		A129_00528	1,442	hypothetical protein	hypothetical protein
16,856		A129_00529	359	hypothetical protein	hypothetical protein
16,856		A129_00530	653	putative HTH-type transcriptional regulator	putative HTH-type transcriptional regulator
16,856		A129_00531	248	hypothetical protein	hypothetical protein
16,856		A129_00532	206	hypothetical protein	hypothetical protein
16,856		A129_00533	1,205	Putative lambdoid prophage Rac integrase	Putative lambdoid prophage Rac integrase
16,856		A129_00534	815	hypothetical protein	hypothetical protein
16,856		A129_00535	791	hypothetical protein	hypothetical protein
16,856		A129_00536	203	hypothetical protein	hypothetical protein

16,856		A129_02668	545	hypothetical protein	hypothetical protein
16,856		A129_02667	836	Integrase	Integrase
16,856		A129_02666	293	transposase, IS911	transposase, IS911
16,856		A129_02665	329	hypothetical protein	hypothetical protein
16,856		A129_02664	1,370	hypothetical protein	hypothetical protein
16,856		A129_02663	1,340	hypothetical protein	hypothetical protein
16,856		A129_02662	1,358	KAP family P-loop domain protein	KAP family P-loop domain protein
16,856		A129_02661	1,064	hypothetical protein	hypothetical protein
16,856		A129_02660	515	acetyltransferase	acetyltransferase
16,856		A129_02659	224	hypothetical protein	hypothetical protein
64,062		A129_00172	587	hypothetical protein	hypothetical protein
64,062		A129_00171	413	putative exported protein	putative exported protein
64,062	csrA-2	A129_00170	200	carbon storage regulator	carbon storage regulator
64,062	lvrB_1	A129_00169	431	Legionella vir region protein	Legionella vir region protein
64,062	lvrA_1	A129_00168	899	LvrA	LvrA
64,062		A129_00167	254	putative phage repressor	putative phage repressor
64,062		A129_00166	851	putative secreted protein	putative secreted protein
64,062		A129_00165	146	hypothetical protein	hypothetical protein
64,062	ctpA_1	A129_00164	2,708	cation efflux transporter	cation efflux transporter
64,062	cebC_1	A129_00163	1,259	chemiosmotic efflux system B protein C	chemiosmotic efflux system B protein C
64,062	cebB_1	A129_00162	1,133	chemiosmotic efflux system B protein B	chemiosmotic efflux system B protein B
64,062	cebA_1	A129_00161	3,143	chemiosmotic efflux system B protein A	chemiosmotic efflux system B protein A
64,062		A129_00160	1,361	metallo-beta lactamase family protein	metallo-beta lactamase family protein
64,062	deoA	A129_00159	1,517	thymidine phosphorylase TdRPase	thymidine phosphorylase TdRPase
64,062		A129_00158	905	ribose-phosphate pyrophosphokinase	ribose-phosphate pyrophosphokinase

64,062		A129_00157	173	hypothetical protein	hypothetical protein
64,062	copA1	A129_00156	2,216	copper efflux ATPase	copper efflux ATPase
64,062	yegE_2	A129_00155	881	sensory box/GGDEF/EAL family protein	sensory box/GGDEF/EAL family protein
64,062		A129_00154	371	putative outer membrane lipoprotein	putative outer membrane lipoprotein
64,062	cecC_2	A129_00153	1,298	chemiosmotic efflux system C protein C	chemiosmotic efflux system C protein C
64,062	cecB	A129_00152	953	chemiosmotic efflux system C protein B	chemiosmotic efflux system C protein B
64,062		A129_00151	3,212	Chemiosmotic efflux system protein A-like protein	Chemiosmotic efflux system protein A-like protein
64,062		A129_00150	452	hypothetical protein	hypothetical protein
64,062		A129_00149	311	hypothetical protein	hypothetical protein
64,062		A129_00148	1,679	hypothetical protein	hypothetical protein
64,062		A129_00147	155	hypothetical protein	hypothetical protein
64,062		A129_00146	527	hypothetical protein	hypothetical protein
64,062	copA2_2	A129_00145	3,086	copper efflux ATPase	copper efflux ATPase
64,062		A129_00144	461	hypothetical protein	hypothetical protein
64,062		A129_00143	611	hypothetical protein	hypothetical protein
64,062		A129_00142	173	hypothetical protein	hypothetical protein
64,062	artJ_2	A129_00141	509	arginine 3rd transport system periplasmic binding protein	arginine 3rd transport system periplasmic binding protein
64,062	helC_2	A129_00140	1,250	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein
64,062	helB_2	A129_00139	1,253	cation efflux system HelB	cation efflux system HelB
64,062	helA_2	A129_00138	3,158	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA
64,062		A129_00137	200	hypothetical protein	hypothetical protein
64,062	nmtR	A129_00136	296	HTH-type transcriptional regulator NmtR	HTH-type transcriptional regulator NmtR
64,062		A129_00135	131	hypothetical protein	hypothetical protein

64,062	cadA_2	A129_00134	2,141	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA
64,062		A129_00133	98	hypothetical protein	hypothetical protein
64,062		A129_00132	173	hypothetical protein	hypothetical protein
64,062	xerC_2	A129_00131	575	Tyrosine recombinase XerC	Tyrosine recombinase XerC
64,062		A129_00130	1,520	hypothetical protein	hypothetical protein
64,062		A129_00129	1,280	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein
64,062		A129_00128	302	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein
64,062		A129_00127	296	hypothetical protein	hypothetical protein
64,062		A129_00126	665	putative S-isoprenylcysteine methyltransferase	putative S-isoprenylcysteine methyltransferase
64,062		A129_00125	500	hypothetical protein	hypothetical protein
64,062		A129_00124	821	hypothetical protein	hypothetical protein
64,062		A129_00123	761	hypothetical protein	hypothetical protein
64,062		A129_00122	1,889	oxidoreductase, FAD-binding	oxidoreductase, FAD-binding
64,062		A129_00121	869	hypothetical protein	hypothetical protein
64,062		A129_00120	1,472	ATP synthase F1, subunit alpha	ATP synthase F1, subunit alpha
64,062	atpF_1	A129_00119	743	ATP synthase subunit b	ATP synthase subunit b
64,062	atpE_1	A129_00118	275	ATP synthase subunit c	ATP synthase subunit c
64,062	atpB_1	A129_00117	692	ATP synthase F0, A subunit	ATP synthase F0, A subunit
64,062		A129_00116	278	Putative F0F1-ATPase subunit (ATPase_gene1)	Putative F0F1-ATPase subunit (ATPase_gene1)
64,062		A129_00115	410	F0F1 ATP synthase subunit epsilon	F0F1 ATP synthase subunit epsilon
64,062	atpD_1	A129_00114	1,424	ATP synthase F1, beta chain	ATP synthase F1, beta chain
64,062		A129_00113	440	hypothetical protein	hypothetical protein
64,062		A129_00112	152	hypothetical protein	hypothetical protein
64,062		A129_00111	263	anaerobic benzoate catabolism transcriptional regulator	anaerobic benzoate catabolism transcriptional regulator

64,062		A129_00110	542	guanylate cyclase	guanylate cyclase
64,062	rre41_1	A129_00109	779	sensory box (GGDEF/EAL domain) regulatory protein	sensory box (GGDEF/EAL domain) regulatory protein
16,989		A129_00097	203	prophage regulatory protein-like protein	prophage regulatory protein-like protein
16,989		A129_00096	965	AbiD phage protein-like protein	AbiD phage protein-like protein
16,989	int_1	A129_00095	1,181	integrase, phage related	integrase, phage related
16,989		A129_00094	275	TnpA transposase	TnpA transposase
16,989		A129_02951	1,103	hypothetical protein	hypothetical protein
16,989		A129_02952	1,184	hypothetical protein	hypothetical protein
16,989		A129_02953	515	hypothetical protein	hypothetical protein
16,989		A129_02954	1,709	inner membrane protein	inner membrane protein
16,989		A129_02955	371	hypothetical protein	hypothetical protein
16,989		A129_02956	152	hypothetical protein	hypothetical protein
16,989		A129_02957	1,355	deoxyguanosinetriphosphate triphosphohydrolase	deoxyguanosinetriphosphate triphosphohydrolase
16,989		A129_02958	1,496	reverse transcriptase	reverse transcriptase
16,989		A129_02959	407	hypothetical protein	hypothetical protein
16,989		A129_02960	683	hypothetical protein	hypothetical protein
16,989		A129_02961	1,031	hypothetical protein	hypothetical protein
16,989		A129_02962	1,244	phage related integrase	phage related integrase
18,290		A129_02121	170	hypothetical protein	
18,290	cadA_4	A129_02120	2,135	cadmium translocating P-type ATPase CadA	
18,290	helA_4	A129_02119	3,149	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	
18,290	helB_4	A129_02118	1,256	cation efflux system HelB	
18,290	helC_5	A129_02117	1,244	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	

18,290		A129_02116	662	phage repressor	
18,290	lvrA_2	A129_02115	872	Legionella vir region protein	
18,290	lvrB_2	A129_02114	383	Legionella vir region protein	
18,290	lvrC	A129_02113	197	Legionella vir region protein	
18,290		A129_02112	434	putative exported protein	
18,290		A129_02111	614	hypothetical protein	
18,290		A129_02110	836	putative exported protein	
18,290		A129_02109	458	hypothetical protein	
18,290		A129_02108	698	hypothetical protein	
18,290		A129_02107	323	hypothetical protein	
18,290		A129_02106	263	hypothetical protein	
18,290		A129_02105	386	hypothetical protein	
18,290		A129_02104	350	hypothetical protein	
18,290		A129_02103	662	hypothetical protein	
18,290		A129_02102	782	hypothetical protein	
18,290		A129_02101	1,313	exported membrane protein	
18,290		A129_02100	389	hypothetical protein	
11,562		A129_02091	209	hypothetical protein	
11,562		A129_02090	968	hypothetical protein	
11,562		A129_02089	803	hypothetical protein	
11,562		A129_02088	899	putative integrase	
11,562		A129_02087	185	hypothetical protein	
11,562		A129_02086	506	antirestriction protein	
11,562		A129_02085	617	hypothetical protein	
11,562		A129_02084	491	hypothetical protein	

11,562	dam_1	A129_02083	818	DNA adenine methylase	
11,562		A129_02899	1,025	Integrase core domain protein	
11,562	tnpA_3	A129_02900	1,418	Transposase for transposon Tn5	
11,562		A129_02901	209	cold shock DNA binding domain protein	
5,373		A129_01064	95	hypothetical protein	hypothetical protein
5,373	int_2	A129_01065	428	integrase, phage related	integrase, phage related
5,373		A129_01066	1,457	Divergent AAA domain protein	Divergent AAA domain protein
5,373		A129_01067	587	hypothetical protein	hypothetical protein
5,373		A129_01068	494	hypothetical protein	hypothetical protein
5,373		A129_01069	677	phage repressor	phage repressor
5,373		A129_01070	206	prophage regulatory protein-like protein	prophage regulatory protein-like protein
5,373		A129_01071	398	LvrA	LvrA
48,997		A129_00623	1,457	Tn3 transposase DDE domain protein	
48,997		A129_00624	965	Tn3 transposase DDE domain protein	
48,997		A129_00625	476	transcriptional regulator, MerR family, mercury resistance	
48,997		A129_00626	452	hypothetical protein	
48,997		A129_00627	584	site specific recombinase	
48,997		A129_00628	506	hypothetical protein	
48,997		A129_02669	305	hypothetical protein	
48,997		A129_02670	209	hypothetical protein	
48,997		A129_02671	425	single strand DNA binding protein	
48,997		A129_02672	224	hypothetical protein	
48,997		A129_02673	167	hypothetical protein	
48,997		A129_02515	275	TnpA transposase	

48,997		A129_00538	227	hypothetical protein	
48,997		A129_02674	290	endonuclease	
48,997		A129_02949	755	Insertion sequence putative ATP-binding protein	
48,997		A129_02950	1,520	Putative transposase for insertion sequence element	
48,997	tufA	A129_02965	1,190	translation elongation factor Tu (EF-Tu)	
48,997		A129_02675	878	Integrase	
48,997		A129_02676	287	transposase, IS911	
48,997		A129_00304	92	hypothetical protein	
48,997		A129_00305	2,621	hypothetical protein	
48,997		A129_00306	1,121	Sid related protein-like protein	
48,997	sdeC	A129_00307	2,045	substrates of the Legionella pneumophila Dot/Icm system SdeC	
12,225		A129_00648	989	hypothetical protein	
12,225		A129_00649	401	Cation efflux system protein cusA	
12,225		A129_00650	362	putative outer membrane lipoprotein	
12,225		A129_00651	404	Response regulator receiver domain protein	
12,225	cadA_3	A129_00652	2,150	cadmium translocating P-type ATPase CadA	
12,225		A129_00653	269	Double zinc ribbon	
12,225	kmtR_2	A129_00654	296	HTH-type transcriptional regulator KmtR	
12,225		A129_00655	206	hypothetical protein	
12,225	helA_3	A129_00656	3,149	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	
12,225	helB_3	A129_00657	1,244	cation efflux system HelB	
12,225	helC_3	A129_00658	863	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	
12,225	helC_4	A129_00659	323	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	
12,225	artJ_4	A129_00660	788	arginine 3rd transport system periplasmic binding protein	

59,011		A129_00008	548	ORF6N domain protein	
59,011	czcD_1	A129_00009	602	cation efflux system protein	
59,011	copA2_1	A129_00010	1,442	copper efflux ATPase	
59,011		A129_00011	449	hypothetical protein	
59,011		A129_00012	3,236	Chemiosmotic efflux system protein A-like protein	
59,011	ceaB_1	A129_00013	971	Chemiosmotic efflux system protein B-like protein	
59,011	cecC_1	A129_00014	1,316	chemiosmotic efflux system C protein C	
59,011		A129_00015	2,198	copper efflux ATPase	
59,011	csoR	A129_00016	275	Copper-sensing transcriptional repressor CsoR	
59,011	yegE_1	A129_00017	839	sensory box/GGDEF/EAL family protein	
59,011		A129_00018	3,002	Tn3 transposase DDE domain protein	
59,011	pinR	A129_00019	575	Putative DNA-invertase from lambdoid prophage Rac	
59,011		A129_00020	791	haloalkane dehalogenase	
59,011		A129_00021	164	hypothetical protein	
59,011		A129_00022	386	Response regulator receiver domain protein	
59,011	cadA_1	A129_00023	2,156	cadmium translocating P-type ATPase CadA	
59,011		A129_00024	269	Double zinc ribbon	
59,011	kmtR_1	A129_00025	296	HTH-type transcriptional regulator KmtR	
59,011	helA_1	A129_00026	3,152	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	
59,011	helB_1	A129_00027	1,244	cation efflux system HelB	
59,011	helC_1	A129_00028	1,247	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	
59,011	artJ_1	A129_00029	788	arginine 3rd transport system periplasmic binding protein	
59,011		A129_00030	188	carbon storage regulator CsrA	
59,011		A129_00031	602	inclusion membrane protein A	

59,011		A129_00032	1,166	hypothetical protein	
59,011		A129_00033	899	hypothetical protein	
59,011		A129_00034	749	hypothetical protein	
59,011		A129_00035	851	hypothetical protein	
59,011		A129_00036	584	transposase (resolvase, DNA invertase)	
59,011		A129_00037	3,050	Tn3 transposase DDE domain protein	
59,011		A129_00038	707	inclusion membrane protein A	
59,011		A129_00039	950	hypothetical protein	
59,011		A129_00040	299	hypothetical protein	
59,011		A129_00041	842	hypothetical protein	
59,011		A129_00042	959	CobQ/CobB/MinD/ParA nucleotide binding domain protein	
59,011		A129_00043	1,151	hypothetical protein	
59,011		A129_00044	464	hypothetical protein	
59,011	parB_1	A129_00045	878	chromosome partitioning protein ParB (SpoOJ)	
59,011		A129_00046	1,043	sulfurylase ThiF	
59,011		A129_00047	1,412	hypothetical protein	
59,011		A129_00048	1,166	hypothetical protein	
59,011		A129_00049	128	hypothetical protein	
59,011		A129_00050	251	hypothetical protein	
59,011		A129_00051	1,034	YcaO-like family protein	
59,011		A129_00052	1,052	TfuA-like protein	
59,011		A129_00053	1,259	hypothetical protein	
59,011		A129_00054	1,010	hypothetical protein	
59,011		A129_00055	533	hypothetical protein	

59,011		A129_00056	176	hypothetical protein	
59,011	csrA_1	A129_00057	227	global regulator (carbon storage regulator)	
59,011	ylpA	A129_00058	752	Lipoprotein YlpA precursor	
59,011		A129_00059	461	putative exported protein	
59,011		A129_00060	647	hypothetical protein	
59,011		A129_00061	845	putative exported protein	
59,011		A129_00062	476	hypothetical protein	
59,011		A129_00063	746	hypothetical protein	
59,011		A129_00064	326	hypothetical protein	
59,011		A129_00065	257	hypothetical protein	
9,013		A129_00090	347	hypothetical protein	
9,013		A129_00091	236	hypothetical protein	
9,013		A129_00092	806	hypothetical protein	
9,013		A129_00093	899	putative integrase	
9,013		A129_01109	365	hypothetical protein	
9,013		A129_01110	209	hypothetical protein	
9,013		A129_01111	425	single strand DNA binding protein	
9,013		A129_01112	224	hypothetical protein	
9,013		A129_01113	167	hypothetical protein	
9,013	tnpA_2	A129_02677	1,418	Transposase for transposon Tn5	
GI: Genomic Island					

Table A8: *L. pneumophila* branch three isolates from HZI genomic islands and their gene products						
GI Length (bp)	**Gene ID**	**Locus**	**Gene Length (bp)**	**Product (H34_Gt22(102)_D)**	**Product (H29_Gt22(100)_D)**	**Product (H35_Gt22(102)_D)**
8,009	thdF	H34_01366	1,334	GTP binding protein in thiophene and furan oxidation (GTPase)	GTP binding protein in thiophene and furan oxidation (GTPase)	GTP binding protein in thiophene and furan oxidation (GTPase)
8,009	oxaA	H34_01365	1,670	Preprotein translocase subunit YidC	Preprotein translocase subunit YidC	Preprotein translocase subunit YidC
8,009	0	H34_01364	227	Putative membrane protein insertion efficiency factor	Putative membrane protein insertion efficiency factor	Putative membrane protein insertion efficiency factor
8,009	rnpA	H34_01363	344	ribonuclease P protein component	ribonuclease P protein component	ribonuclease P protein component
8,009	rpmH	H34_01362	134	50S ribosomal protein L34	50S ribosomal protein L34	50S ribosomal protein L34
8,009	dnaA	H34_01361	1,358	chromosomal replication initiator protein DnaA	chromosomal replication initiator protein DnaA	chromosomal replication initiator protein DnaA
8,009	dnaN	H34_01360	1,103	DNA polymerase III beta chain	DNA polymerase III beta chain	DNA polymerase III beta chain
8,009	recF	H34_01359	1,061	DNA recombination and repair protein ATPase RecF	DNA recombination and repair protein ATPase RecF	DNA recombination and repair protein ATPase RecF
16,952	0	H34_01220	539	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01219	275	Prophage CP4-57 regulatory protein alpA	Prophage CP4-57 regulatory protein alpA	Prophage CP4-57 regulatory protein alpA
16,952	0	H34_01218	281	DNA primase	DNA primase	DNA primase
16,952	0	H34_01217	2,084	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01216	437	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01215	680	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01214	299	Helix-turn-helix domain protein	Helix-turn-helix domain protein	Helix-turn-helix domain protein
16,952	0	H34_01213	2,477	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01212	1,298	Competence protein CoiA-like family protein	Competence protein CoiA-like family protein	Competence protein CoiA-like family protein
16,952	0	H34_01211	632	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01210	548	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01209	710	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01208	806	Calcineurin-like phosphoesterase	Calcineurin-like phosphoesterase	Calcineurin-like phosphoesterase

16,952	0	H34_01207	662	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01206	917	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01205	755	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01204	371	hypothetical protein	hypothetical protein	hypothetical protein
16,952	0	H34_01203	506	Acetyltransferase (GNAT) family protein	Acetyltransferase (GNAT) family protein	Acetyltransferase (GNAT) family protein
16,952	yfkM	H34_01202	674	General stress protein 18	General stress protein 18	General stress protein 18
7,152	0	H34_02053	992	multidrug resistance secretion protein	multidrug resistance secretion protein	multidrug resistance secretion protein
7,152	0	H34_02052	398	hypothetical protein	hypothetical protein	hypothetical protein
7,152	0	H34_02051	638	bacteriophage related DNA polymerase	bacteriophage related DNA polymerase	bacteriophage related DNA polymerase
7,152	0	H34_02050	242	MFS transporter family protein	MFS transporter family protein	MFS transporter family protein
7,152	0	H34_02049	995	MFS transporter family protein	MFS transporter family protein	MFS transporter family protein
7,152	0	H34_02048	656	N-acetylmuramoyl-L-alanine amidase; amidase 2	N-acetylmuramoyl-L-alanine amidase; amidase 2	N-acetylmuramoyl-L-alanine amidase; amidase 2
7,152	copA2_2	H34_02047	1,541	copper efflux ATPase	copper efflux ATPase	copper efflux ATPase
7,152	0	H34_02046	434	hypothetical protein	hypothetical protein	hypothetical protein
27,517	0	H34_00368	587	hypothetical protein	hypothetical protein	hypothetical protein
27,517	0	H34_00369	413	putative exported protein	0	putative exported protein
27,517	csrA-2	H34_00370	200	carbon storage regulator	0	carbon storage regulator
27,517	lvrB_1	H34_00371	431	Legionella vir region protein	0	Legionella vir region protein
27,517	lvrA_1	H34_00372	899	LvrA	0	LvrA
27,517	0	H34_00373	254	putative phage repressor	putative phage repressor	putative phage repressor
27,517	0	H34_00374	851	putative secreted protein	putative secreted protein	putative secreted protein
27,517	0	H34_00375	146	hypothetical protein	hypothetical protein	hypothetical protein
27,517	ctpA_1	H34_00376	2,705	cation efflux transporter	cation efflux transporter	cation efflux transporter
27,517	cebC_1	H34_00377	1,259	chemiosmotic efflux system B protein C	chemiosmotic efflux system B protein C	chemiosmotic efflux system B protein C
27,517	cebB_1	H34_00378	1,133	chemiosmotic efflux system B protein B	chemiosmotic efflux system B protein B	chemiosmotic efflux system B protein B

27,517	cebA	H34_00379	3,143	chemiosmotic efflux system B protein A	chemiosmotic efflux system B protein A	chemiosmotic efflux system B protein A
27,517	0	H34_00380	1,361	metallo-beta lactamase family protein	metallo-beta lactamase family protein	metallo-beta lactamase family protein
27,517	deoA	H34_00381	1,517	thymidine phosphorylase TdRPase	thymidine phosphorylase TdRPase	thymidine phosphorylase TdRPase
27,517	0	H34_00382	932	ribose-phosphate pyrophosphokinase	ribose-phosphate pyrophosphokinase	ribose-phosphate pyrophosphokinase
27,517	0	H34_00383	167	hypothetical protein	hypothetical protein	hypothetical protein
27,517	copA1_1	H34_00384	2,216	copper efflux ATPase	copper efflux ATPase	copper efflux ATPase
27,517	yegE	H34_00385	881	sensory box/GGDEF/EAL family protein	sensory box/GGDEF/EAL family protein	sensory box/GGDEF/EAL family protein
27,517	0	H34_00386	371	putative outer membrane lipoprotein	putative outer membrane lipoprotein	putative outer membrane lipoprotein
27,517	cecC	H34_00387	1,298	chemiosmotic efflux system C protein C	chemiosmotic efflux system C protein C	chemiosmotic efflux system C protein C
27,517	cecB	H34_00388	971	chemiosmotic efflux system C protein B	chemiosmotic efflux system C protein B	chemiosmotic efflux system C protein B
27,517	0	H34_00389	3,212	Chemiosmotic efflux system protein A-like protein	Chemiosmotic efflux system protein A-like protein	Chemiosmotic efflux system protein A-like protein
27,517	0	H34_00390	452	hypothetical protein	hypothetical protein	hypothetical protein
27,517	0	H34_00391	1,055	Leucine carboxyl methyltransferase	Leucine carboxyl methyltransferase	Leucine carboxyl methyltransferase
32779	0	H34_00405	173	hypothetical protein	hypothetical protein	hypothetical protein
32779	artJ_1	H34_00406	509	arginine 3rd transport system periplasmic binding protein	arginine 3rd transport system periplasmic binding protein	arginine 3rd transport system periplasmic binding protein
32779	helC_1	H34_00407	1,250	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein
32779	helB_1	H34_00408	1,253	cation efflux system HelB	cation efflux system HelB	cation efflux system HelB
32779	helA_1	H34_00409	3,158	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA
32779	0	H34_00410	200	hypothetical protein	hypothetical protein	hypothetical protein
32779	nmtR	H34_00411	296	HTH-type transcriptional regulator NmtR	HTH-type transcriptional regulator NmtR	HTH-type transcriptional regulator NmtR
32779	0	H34_00412	131	hypothetical protein	hypothetical protein	hypothetical protein
32779	cadA_1	H34_00413	2,141	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA
32779	0	H34_00414	98	hypothetical protein	hypothetical protein	hypothetical protein

32779	0	H34_00415	173	hypothetical protein	hypothetical protein	hypothetical protein
32779	xerC	H34_00416	575	Tyrosine recombinase XerC	Tyrosine recombinase XerC	Tyrosine recombinase XerC
32779	0	H34_00417	1,520	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00418	1,280	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein
32779	0	H34_00419	302	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein	Tn3 transposase DDE domain protein
32779	0	H34_00420	296	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00421	665	putative S-isoprenylcysteine methyltransferase	putative S-isoprenylcysteine methyltransferase	putative S-isoprenylcysteine methyltransferase
32779	0	H34_00422	500	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00423	821	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00424	761	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00425	1,883	oxidoreductase, FAD-binding	oxidoreductase, FAD-binding	oxidoreductase, FAD-binding
32779	0	H34_00426	374	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00427	869	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00428	1,472	ATP synthase F1, subunit alpha	ATP synthase F1, subunit alpha	ATP synthase F1, subunit alpha
32779	atpF_1	H34_00429	743	ATP synthase subunit b	ATP synthase subunit b	ATP synthase subunit b
32779	atpE_1	H34_00430	275	ATP synthase subunit c	ATP synthase subunit c	ATP synthase subunit c
32779	atpB_1	H34_00431	692	ATP synthase F0, A subunit	ATP synthase F0, A subunit	ATP synthase F0, A subunit
32779	0	H34_00432	278	Putative F0F1-ATPase subunit (ATPase_gene1)	Putative F0F1-ATPase subunit (ATPase_gene1)	Putative F0F1-ATPase subunit (ATPase_gene1)
32779	0	H34_00433	410	F0F1 ATP synthase subunit epsilon	F0F1 ATP synthase subunit epsilon	F0F1 ATP synthase subunit epsilon
32779	atpD_1	H34_00434	1,424	ATP synthase F1, beta chain	ATP synthase F1, beta chain	ATP synthase F1, beta chain
32779	0	H34_00435	449	hypothetical protein	hypothetical protein	hypothetical protein
32779	0	H34_00436	263	anaerobic benzoate catabolism transcriptional regulator	anaerobic benzoate catabolism transcriptional regulator	anaerobic benzoate catabolism transcriptional regulator
32779	0	H34_00437	542	guanylate cyclase	guanylate cyclase	guanylate cyclase
32779	0	H34_00438	353	putative diguanylate cyclase	putative diguanylate cyclase	putative diguanylate cyclase

32779	0	H34_00439	1,163	inner membrane protein PLUS sensory box protein LssE	inner membrane protein PLUS sensory box protein LssE	inner membrane protein PLUS sensory box protein LssE
32779	phbC_1	H34_00440	1,793	polyhydroxyalkanoic synthase	polyhydroxyalkanoic synthase	polyhydroxyalkanoic synthase
16619	0	H34_02537	1,034	transposase (ISSod13)	transposase (ISSod13)	transposase (ISSod13)
16619	0	H34_02538	1,529	TcdA/TcdB catalytic glycosyltransferase domain protein	TcdA/TcdB catalytic glycosyltransferase domain protein	TcdA/TcdB catalytic glycosyltransferase domain protein
16619	0	H34_02539	566	peptidyl-prolyl cis-trans isomerase (rotamase)	peptidyl-prolyl cis-trans isomerase (rotamase)	peptidyl-prolyl cis-trans isomerase (rotamase)
16619	0	H34_02540	2,096	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02541	2,933	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02542	1,565	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02543	227	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02544	113	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02545	560	transcriptional regulator, TetR family	transcriptional regulator, TetR family	transcriptional regulator, TetR family
16619	0	H34_02546	749	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02547	242	hypothetical protein	hypothetical protein	hypothetical protein
16619	0	H34_02548	1,793	Tpr	Tpr	Tpr
16619	gst_1	H34_02549	623	glutathione S-transferase	glutathione S-transferase	glutathione S-transferase
21177	0	H34_01023	395	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01024	1,907	cadmium efflux ATPase	cadmium efflux ATPase	cadmium efflux ATPase
21177	0	H34_01025	170	hypothetical protein	hypothetical protein	hypothetical protein
21177	cadA_2	H34_01026	2,135	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA
21177	helA_2	H34_01027	3,149	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA
21177	helB_2	H34_01028	1,256	cation efflux system HelB	cation efflux system HelB	cation efflux system HelB
21177	helC_2	H34_01029	1,244	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein
21177	0	H34_01030	662	phage repressor	phage repressor	phage repressor

21177	lvrA_2	H34_01031	872	Legionella vir region protein	Legionella vir region protein	Legionella vir region protein
21177	lvrB_2	H34_01032	383	Legionella vir region protein	Legionella vir region protein	Legionella vir region protein
21177	lvrC	H34_01033	197	Legionella vir region protein	Legionella vir region protein	Legionella vir region protein
21177	0	H34_01034	434	putative exported protein	putative exported protein	putative exported protein
21177	0	H34_01035	596	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01036	836	putative exported protein	putative exported protein	putative exported protein
21177	0	H34_01037	458	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01038	698	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01039	323	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01040	263	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01041	386	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01042	350	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01043	662	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01044	782	hypothetical protein	hypothetical protein	hypothetical protein
21177	0	H34_01045	1,313	exported membrane protein	exported membrane protein	exported membrane protein
21177	0	H34_01046	389	hypothetical protein	hypothetical protein	hypothetical protein
30124	0	H34_02359	965	major outer membrane protein	heme exporter protein CcmC	
30124	tufA_2	H34_02301	917	translation elongation factor Tu (EF-Tu)	heme exporter protein CcmD	
30124	0	H34_02302	341	preprotein translocase subunit SecE	cytochrome c-type biogenesis protein CcmE	
30124	nusG	H34_02303	548	transcription antitermination protein NusG	cytochrome c-type biogenesis protein CcmF	
30124	rplK	H34_02304	434	50S ribosomal protein L11		
30124	rplA	H34_02305	695	50S ribosomal protein L1		
30124	rplJ	H34_02306	533	50S ribosomal protein L10		
30124	rplL	H34_02307	380	50S ribosomal protein L7/L12	Transposase DDE domain protein	
30124	rpoB	H34_02308	4,106	DNA-directed RNA polymerase beta subunit	hypothetical protein	

30124	rpoC	H34_02309	4,247	DNA-directed RNA polymerase subunit beta'	transposase (ISSod13)	
30124	rpsL	H34_02310	380	30S ribosomal protein S12	Transposase insI for insertion sequence element IS30B/C/D	
30124	rpsG	H34_02311	527	30S ribosomal protein S7		
30124	fusA	H34_02312	2,084	translation elongation factor G (EF-G)		
30124	rpoD	H34_02405	1,865	RNA polymerase sigma 70 factor (RpoD)	RNA polymerase sigma 70 factor (RpoD)	RNA polymerase sigma 70 factor (RpoD)
30124		H34_02407	776	hypothetical protein	hypothetical protein	hypothetical protein
30124	recD2	H34_02408	2,651	ATP-dependent RecD-like DNA helicase	ATP-dependent RecD-like DNA helicase	ATP-dependent RecD-like DNA helicase
30124	0	H34_02409	1,487	reverse transcriptase	reverse transcriptase	reverse transcriptase
30124	0	H34_02410	206	hypothetical protein	hypothetical protein	hypothetical protein
30124	0	H34_02411	263	Conjugal transfer protein TraD	Conjugal transfer protein TraD	Conjugal transfer protein TraD
30124	0	H34_02412	350	putative RNA helicase	putative RNA helicase	putative RNA helicase

HZI: Helmholtz Center for Infection Research GI: Genomic Island

Table A9: *L. pneumophila* branch six sub-lineage isolates from the West bank and Germany genomic islands and their gene products

GI Length (bp)	Gene ID	Locus	Gene Length (bp)	Product (A193_Gt40(47)_Ps)	Product (A195_Gt40(47)_Ps)	Product (L04-545_Cl_Gt40(47)_D)
6,667	oxaA	A193_01430	1,670	Prep+K2:M74rotein translocase subunit YidC	Preprotein translocase subunit YidC	Preprotein translocase subunit YidC
6,667		A193_01431	227	Putative membrane protein insertion efficiency factor	Putative membrane protein insertion efficiency factor	Putative membrane protein insertion efficiency factor
6,667	rnpA	A193_01432	344	ribonuclease P protein component	ribonuclease P protein component	ribonuclease P protein component
6,667	rpmH	A193_01433	134	50S ribosomal protein L34	50S ribosomal protein L34	50S ribosomal protein L34
6,667	dnaA	A193_01434	1,358	chromosomal replication initiator protein DnaA	chromosomal replication initiator protein DnaA	chromosomal replication initiator protein DnaA
6,667	dnaN	A193_01435	1,103	DNA polymerase III beta chain	DNA polymerase III beta chain	DNA polymerase III beta chain
6,667	recF_1	A193_01436	1,061	DNA recombination and repair protein ATPase RecF	DNA recombination and repair protein ATPase RecF	DNA recombination and repair protein ATPase RecF
19,317		A193_02966	365	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02967	266	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02968	326	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02969	632	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02970	431	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02971	821	putative exported protein	putative exported protein	hypothetical protein
19,317		A193_02972	587	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02973	413	putative exported protein	putative exported protein	hypothetical protein
19,317	csrA-2	A193_02974	200	carbon storage regulator	carbon storage regulator	putative exported protein
19,317	lvrB	A193_02975	362	Legionella vir region protein	Legionella vir region protein	hypothetical protein
19,317	lvrA_2	A193_02976	899	LvrA	LvrA	putative exported protein
19,317		A193_02977	254	putative phage repressor	putative phage repressor	carbon storage regulator
19,317	copA2_2	A193_02978	482	copper efflux ATPase	copper efflux ATPase	Legionella vir region protein
19,317		A193_02979	461	putative outer membrane lipoprotein	putative outer membrane lipoprotein	LvrA
19,317		A193_02980	611	hypothetical protein	hypothetical protein	putative phage repressor

19,317		A193_02981	173	hypothetical protein	hypothetical protein	copper efflux ATPase
19,317	artJ_4	A193_02982	329	arginine 3rd transport system periplasmic binding protein	arginine 3rd transport system periplasmic binding protein	putative outer membrane lipoprotein
19,317	helC	A193_02983	1,250	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein	hypothetical protein
19,317	helB	A193_02984	1,253	cation efflux system HelB	cation efflux system HelB	hypothetical protein
19,317	helA	A193_02985	3,158	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA	arginine 3rd transport system periplasmic binding protein
19,317		A193_02986	200	hypothetical protein	hypothetical protein	cobalt/zinc/cadmium efflux RND transporter, outer membrane protein
19,317	nmtR	A193_02987	296	HTH-type transcriptional regulator NmtR	HTH-type transcriptional regulator NmtR	cation efflux system HelB
19,317		A193_02988	131	hypothetical protein	hypothetical protein	cobalt/zinc/cadmium efflux RND transporter, permease protein HelA
19,317	cadA	A193_02989	2,141	cadmium translocating P-type ATPase CadA	cadmium translocating P-type ATPase CadA	hypothetical protein
19,317		A193_02990	167	hypothetical protein	hypothetical protein	HTH-type transcriptional regulator NmtR
19,317		A193_02991	173	hypothetical protein	hypothetical protein	hypothetical protein
19,317		A193_02992	254	anaerobic benzoate catabolism transcriptional regulator	anaerobic benzoate catabolism transcriptional regulator	cadmium translocating P-type ATPase CadA
17,424		A193_03004	245	hypothetical protein	hypothetical protein	hypothetical protein
17,424		A193_03005	218	prophage regulatory protein-like protein	prophage regulatory protein-like protein	hypothetical protein
17,424		A193_03006	1,244	integrase, phage related	integrase, phage related	anaerobic benzoate catabolism transcriptional regulator
17,424		A193_03007	650	hypothetical protein	hypothetical protein	hypothetical protein
17,424		A193_03008	2,093	hypothetical protein	hypothetical protein	prophage regulatory protein-like protein
17,424		A193_03009	272	hypothetical protein	hypothetical protein	integrase, phage related
17,424		A193_03010	1,751	inner membrane protein	inner membrane protein	hypothetical protein
17,424		A193_03011	809	hypothetical protein	hypothetical protein	hypothetical protein
17,424	recF_2	A193_03012	1,985	DNA replication and repair protein RecF	DNA replication and repair protein RecF	hypothetical protein
17,424	uvrD_2	A193_03013	1,856	DNA dependent ATPase I and helicase II	DNA dependent ATPase I and helicase II	inner membrane protein

17,424	dinJ_2	A193_03014	278	Antitoxin DinJ	Antitoxin DinJ	hypothetical protein
17,424		A193_03015	359	hypothetical protein	hypothetical protein	DNA replication and repair protein RecF
17,424		A193_03016	143	hypothetical protein	Helix-turn-helix domain protein	DNA dependent ATPase I and helicase II
17,424		A193_03017	446	Helix-turn-helix domain protein	hypothetical protein	Antitoxin DinJ
17,424		A193_03018	575	hypothetical protein	putative regulator PrlF	hypothetical protein
17,424		A193_03019	218	putative regulator PrlF	oxidoreductase (L-gululonolactone oxidase)	hypothetical protein
17,424		A193_03020	1,250	phage related integrase	oxidoreductase, short chain dehydrogenase/reductase family protein	Helix-turn-helix domain protein
7,119		A193_00072	1,295	oxidoreductase (L-gululonolactone oxidase)	transcriptional regulator, LysR family	hypothetical protein
7,119		A193_00073	722	oxidoreductase, short chain dehydrogenase/reductase family protein	Transposase	putative regulator PrlF
7,119		A193_00074	926	transcriptional regulator, LysR family	hypothetical protein	oxidoreductase (L-gululonolactone oxidase)
7,119		A193_00075	233	Transposase	hypothetical protein	oxidoreductase, short chain dehydrogenase/reductase family protein
7,119		A193_00076	1,241	hypothetical protein	hypothetical protein	transcriptional regulator, LysR family
7,119		A193_00077	1,196	hypothetical protein	pyridoxamine 5'-phosphate oxidase	Transposase
7,119		A193_00078	197	hypothetical protein	putative transcriptional regulator, MarR family	hypothetical protein
68,131	pdxH_2	A193_00453	587	pyridoxamine 5'-phosphate oxidase	glucose-1-dehydrogenase, (3-oxoacyl-(acyl carrier protein) reductase)	hypothetical protein
68,131		A193_00452	380	putative transcriptional regulator, MarR family	hypothetical protein	hypothetical protein
68,131		A193_00451	764	glucose-1-dehydrogenase, (3-oxoacyl-(acyl carrier protein) reductase)	hypothetical protein	pyridoxamine 5'-phosphate oxidase
68,131		A193_00450	158	hypothetical protein	F-box protein	putative transcriptional regulator, MarR family
68,131		A193_00449	941	hypothetical protein	two component response regulator, CheY-like receiver domain	glucose-1-dehydrogenase, (3-oxoacyl-(acyl carrier protein) reductase)
68,131	pof1	A193_00448	518	F-box protein	response regulator TutC	hypothetical protein
68,131		A193_00447	419	two component response regulator, CheY-like receiver domain	hypothetical protein	hypothetical protein
68,131	stuC	A193_00446	2,450	response regulator TutC	hypothetical protein	F-box protein

68,131		A193_00445	1,448	hypothetical protein	hypothetical protein	two component response regulator, CheY-like receiver domain
68,131		A193_00444	1,178	hypothetical protein	hypothetical protein	response regulator TutC
68,131		A193_00443	359	hypothetical protein	aminoglycoside 6-adenylyltransferase	hypothetical protein
68,131		A193_00442	578	hypothetical protein	multidrug resistance ABC transporter ATP-binding protein	hypothetical protein
68,131		A193_00441	866	aminoglycoside 6-adenylyltransferase	substrates of the Legionella pneumophila Dot/Icm system SdeC	hypothetical protein
68,131	abcT3_1	A193_00440	1,796	multidrug resistance ABC transporter ATP-binding protein	Sid related protein-like protein	hypothetical protein
68,131	sdeC	A193_00439	4,607	substrates of the Legionella pneumophila Dot/Icm system SdeC	hypothetical protein	aminoglycoside 6-adenylyltransferase
68,131		A193_00438	1,121	Sid related protein-like protein	substrates of the Legionella pneumophila Dot/Icm system SdeB	multidrug resistance ABC transporter ATP-binding protein
68,131		A193_00437	2,627	hypothetical protein	ankyrin repeat family protein	substrates of the Legionella pneumophila Dot/Icm system SdeB
68,131	sdeB_2	A193_00436	5,765	substrates of the Legionella pneumophila Dot/Icm system SdeB	hypothetical protein	ankyrin repeat family protein
68,131		A193_00435	2,267	ankyrin repeat family protein	hypothetical protein	hypothetical protein
68,131		A193_00434	992	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00433	206	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00432	13,811	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00431	104	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00430	359	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00429	560	CrtF (hydroxyneurosporene (C20) methyltransferase)	CrtF (hydroxyneurosporene (C20) methyltransferase)	CrtF (hydroxyneurosporene (C20) methyltransferase)
68,131	sdeB_1	A193_00428	2,468	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_00427	107	Alpha/beta hydrolase family protein	Alpha/beta hydrolase family protein	Alpha/beta hydrolase family protein
68,131		A193_01961	1745	Glycosyltransferase sugar-binding region containing DXD motif protein	Glycosyltransferase sugar-binding region containing DXD motif protein	Glycosyltransferase sugar-binding region containing DXD motif protein
68,131		A193_01960	623	hypothetical protein	hypothetical protein	hypothetical protein

68,131		A193_01959	1,091	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_01958	1,367	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_01957	2,258	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_01956	848	DNA primase DnaG	DNA primase DnaG	DNA primase DnaG
68,131		A193_01955	428	RNA polymerase sigma 70 factor (RpoD)	RNA polymerase sigma 70 factor (RpoD)	RNA polymerase sigma 70 factor (RpoD)
68,131		A193_01954	512	hypothetical protein	hypothetical protein	hypothetical protein
68,131		A193_01953	1,259	ATP-dependent RecD-like DNA helicase	ATP-dependent RecD-like DNA helicase	ATP-dependent RecD-like DNA helicase
28,258		A193_01759	443	Acetyltransferase (GNAT) family protein	Acetyltransferase (GNAT) family protein	Acetyltransferase (GNAT) family protein
28,258	dnaG	A193_01758	1,733	hypothetical protein	hypothetical protein	hypothetical protein
28,258	rpoD	A193_01757	1,865	Conjugal transfer protein TraD	Conjugal transfer protein TraD	Conjugal transfer protein TraD
28,258		A193_01755	776	hypothetical protein	hypothetical protein	hypothetical protein
28,258	recD2	A193_01754	2,651	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01753	251	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01752	206	inner membrane protein	inner membrane protein	inner membrane protein
28,258		A193_01751	266	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01750	371	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01749	572	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01748	1,655	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01747	1,616	phage related integrase	phage related integrase	phage related integrase
28,258		A193_01746	275	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01745	857	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01744	608	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01743	233	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01742	1,232	23S rRNA (uracil-C(5))-methyltransferase RlmCD	23S rRNA (uracil-C(5))-methyltransferase RlmCD	23S rRNA (uracil-C(5))-methyltransferase RlmCD
28,258		A193_01964	185	SidC protein (substrate of the Dot/Icm system)	SidC protein (substrate of the Dot/Icm system)	SidC protein (substrate of the Dot/Icm system)

28,258		A193_01965	914	transposase (IS652)	transposase (IS652)	transposase (IS652)
28,258		A193_01966	698	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01967	932	hypothetical protein	hypothetical protein	hypothetical protein
28,258	rlmCD	A193_01968	638	putative protein kinase	hypothetical protein	hypothetical protein
28,258	sidC_2	A193_01969	1,694	hypothetical protein	putative integrase	putative integrase
28,258		A193_01970	1,175	hypothetical protein	hypothetical protein	hypothetical protein
28,258		A193_01971	770	Putative type I restriction enzyme R protein	Putative type I restriction enzyme R protein	Putative type I restriction enzyme R protein
19,749		A193_02172	842	hypothetical protein	hypothetical protein	hypothetical protein
19,749		A193_02173	1,388	Putative type-1 restriction enzyme specificity protein MPN_089	Putative type-1 restriction enzyme specificity protein MPN_089	Putative type-1 restriction enzyme specificity protein MPN_089
19,749		A193_02174	683	Putative type I restriction enzyme HindVIIP M protein	Putative type I restriction enzyme HindVIIP M protein	Putative type I restriction enzyme HindVIIP M protein
19,749	cbpA_2	A193_02175	1,700	putative, uroporphyrin-III C- methyltransferase	putative, uroporphyrin-III C- methyltransferase	putative, uroporphyrin-III C- methyltransferase
19,749		A193_02176	389	sporulation initiation inhibitor protein Soj	sporulation initiation inhibitor protein Soj	sporulation initiation inhibitor protein Soj
19,749		A193_02177	179	chromosome partitioning protein ParB (SpoOJ)	chromosome partitioning protein ParB (SpoOJ)	chromosome partitioning protein ParB (SpoOJ)
19,749		A193_02178	206	hypothetical protein	hypothetical protein	hypothetical protein
19,749		A193_02179	524	ISSod6 transposase_	ISSod6 transposase_	ISSod6 transposase_
19,749		A193_02180	551	ISSod6 transposase_	ISSod6 transposase_	ISSod6 transposase_
19,749		A193_02181	833	hypothetical protein	hypothetical protein	hypothetical protein
19,749		A193_02182	335	hypothetical protein	hypothetical protein	hypothetical protein
19,749		A193_02183	377	Cytochrome c oxidase subunit 3	Cytochrome c oxidase subunit 3	Cytochrome c oxidase subunit 3
19,749		A193_02184	896	Cytochrome c oxidase assembly protein CtaG	Cytochrome c oxidase assembly protein CtaG	Cytochrome c oxidase assembly protein CtaG
19,749		A193_02185	125	Cytochrome c oxidase subunit 1	Cytochrome c oxidase subunit 1	Cytochrome c oxidase subunit 1
19,749		A193_02186	3,149	cytochrome c oxidase, subunit II	cytochrome c oxidase, subunit II	cytochrome c oxidase, subunit II
19,749		A193_02187	173	cytochrome c	cytochrome c	cytochrome c
19,749		A193_02188	1,217	Naphthalene 1,2-dioxygenase system ferredoxin subunit	Naphthalene 1,2-dioxygenase system ferredoxin subunit	Naphthalene 1,2-dioxygenase system ferredoxin subunit

19,749		A193_02189	1,571	CapM protein, capsular polysaccharide biosynthesis	CapM protein, capsular polysaccharide biosynthesis	CapM protein, capsular polysaccharide biosynthesis
19,749		A193_02191	377	transporter, LysE family	transporter, LysE family	transporter, LysE family
18,487	parA	A193_02503	770	hypothetical protein	hypothetical protein	hypothetical protein
18,487	parB_2	A193_02504	899	ubiquinone/menaquinone biosynthesis methyltransferase UbiE	ubiquinone/menaquinone biosynthesis methyltransferase UbiE	ubiquinone/menaquinone biosynthesis methyltransferase UbiE
18,487		A193_02505	335	hypothetical protein	hypothetical protein	hypothetical protein
18,487		A193_02506	239	hypothetical protein	hypothetical protein	hypothetical protein

Table A10. List of *L. pneumophila* strains (n=180) isolated from the West Bank analysed in this study, MLVA-8(12) done by (97).

Strain	Sgp, mAb	ST	Lpn PCR	MLVA-8(12)	MLVA-8(12) profile	MLVA_CC	Location	Sample type	Year
A1	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A10	6	1326	+	Gt13(106)	8,8,11,1,4,16,1,1,3,3,6,4,8	VACC2	AQU	Biofilm swab	2013
A100	1	NA	+	Gt38(109)	8,8,6,2,4,8,1,1,3,2,6,4,8	VACC2	Hospital F	Biofilm swab	2013
A101	1	NA	+	Gt63(83)	7,7,10,2,4,0,4,0,0,3,14,4,9	VACC1	Hospital H	Biofilm swab	2013
A102	1	NA	+	Gt63(83)	7,7,10,2,4,0,4,0,0,3,14,4,9	VACC1	Hospital B	Biofilm swab	2013
A103	1	NA	+	Gt63(83)	7,7,10,2,4,0,4,0,0,3,14,4,9	VACC1	Hospital H	Biofilm swab	2013
A104	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A105	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A106	1	1	+	Gt6(15)	7,7,10,2,4,9,4,2,18,2,14,5,5	VACC1	Hospital G	Biofilm swab	2013
A107	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2013
A108	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2013
A109	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital F	Biofilm swab	2013
A110	6 Dresden	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Biofilm swab	2013
A112	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2013
A114	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Water sample	2013
A115	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2013
A116	6 Dresden	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Biofilm swab	2013
A119	6 Dresden	9	+	Gt64(74)	8,8,11,2,4,16,1,1,3,3,6,4,8	VACC2	Hospital F	Biofilm swab	2012
A12	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	AQU	Biofilm swab	2012
A121	6 Dresden	NA	+	Gt55(94)	9,8,6,2,4,13,3,0,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A122	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A123	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm	2013
A124	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A127	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A128	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012

A129	6 Dresden	9	+	Gt64(74)	8,8,11,2,4,16,1,1,3,3,6,4,8	VACC2	Hospital F	Biofilm swab	2012
A13	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	AQU	Biofilm swab	2012
A130	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A131	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A132	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A133	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A134	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A135	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A137	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A138	6 Dresden	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2012
A139	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A14	8	1358	+	Gt12(84)	7,7,9,2,4,17,1,1,18,0,14,5,8		AQU	Biofilm swab	2012
A140	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A141	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A142	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital E	Biofilm swab	2012
A143	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital E	Biofilm swab	2012
A144	(2-14)	1326	+	Gt13(72)	8,8,11,2,4,16,1,1,3,2,6,4,8	VACC2	Hospital E	Biofilm swab	2012
A145	1	1		Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A148	6 Dresden	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Water sample	2012
A149	6 Dresden	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Biofilm swab	2012
A15	8	1358	+	Gt12(84)	7,7,9,2,4,17,1,1,18,0,14,5,8		AQU	Biofilm swab	2012
A152	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A153	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A154	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital F	Biofilm swab	2012
A156	6 Dresden	9	+	Gt64(74)	8,8,11,2,4,16,1,1,3,3,6,4,8	VACC2	Hospital F	Biofilm swab	2013
A157	6 Dresden	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Biofilm swab	2014
A159	6	461	+	Gt10(141)	9,8,6,2,4,13,3,4,10,3,10,0,8	VACC11	Hospital F	Biofilm	2014

	Dresden							swab	
A16	8	1358	+	Gt12(84)	7,7,9,2,4,17,1,1,18,0,14,5,8		AQU	Biofilm swab	2012
A161	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital F	Biofilm swab	2012
A162	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A163	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A164	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A165	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A166	(2-14)	1482	+	Gt8(142)	10,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital G	Biofilm swab	2012
A167	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A168	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A169	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A17	8	1358	+	Gt12(84)	7,7,9,2,4,17,1,1,18,0,14,5,8		AQU	Biofilm swab	2012
A170	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A171	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A172	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A173	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A174	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A175	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital C	Biofilm swab	2013
A176	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital C	Biofilm swab	2014
A177	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2014
A178	6 Dresden	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2014
A179	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm	2014
A18	8	1358	+	Gt12(84)	7,7,9,2,4,17,1,1,18,0,14,5,8		AQU	Biofilm swab	2014
A180	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A181	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital F	Biofilm swab	2014
A182	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2014
A183	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2014
A184	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014

A186	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A187	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A188	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A189	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A19	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A190	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A191	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A192	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2014
A193	6 Dresden	292	+	Gt40(47)	8,8,9,2,4,13,2,2,21,3,10,4,8	VACC5	Hospital H	Biofilm swab	2012
A194	6 Dresden	292	+	Gt40(47)	8,8,9,2,4,13,2,2,21,3,10,4,8	VACC5	Hospital H	Biofilm swab	2012
A195	6 Dresden	292	+	Gt40(47)	8,8,9,2,4,13,2,2,21,3,10,4,8	VACC5	Hospital H	Biofilm swab	2012
A196	(2-14)	1326	+	Gt13(143)	8,8,11,1,4,16,1,1,3,0,6,4,8	VACC2	Hospital E	Biofilm swab	2012
A197	10	1326	+	Gt13(143)	8,8,11,1,4,16,1,1,3,0,6,4,8	VACC2	Hospital E	Biofilm swab	2012
A198	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A2	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A20	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital F	Biofilm swab	2012
A21	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Water sample	2012
A22	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital H	Biofilm swab	2012
A23	(2-14)	1438	+	Gt16(1)	9,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital G	Biofilm swab	2012
A24	(2-14)	1482	+	Gt8(23)	10,8,8,2,4,13,2,2,18,3,14,4,8	VACC5	Hospital G	Biofilm swab	2012
A25	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A26	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A27	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A28	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A29	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A3	1 OLDA	1	+	Gt4(20)	7,7,10,2,4,9,4,2,17,3,14,4,5	VACC1	AQU	Biofilm swab	2012
A30	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012

A31	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A32	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Water sample	2012
A33	1	1	+	Gt6(18)	7,7,10,2,4,9,4,2,18,3,14,5,5	VACC1	Hospital G	Biofilm swab	2012
A34	(2-14)	1438	+	Gt16(1)	9,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital A	Biofilm swab	2012
A35	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A36	(2-14)	1438	+	Gt16(6)	9,8,8,2,4,13,2,2,18,2,10,4,8	VACC5	Hospital A	Biofilm swab	2012
A37	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital A	Biofilm swab	2012
A38	(2-14)	1438	+	Gt16(1)	9,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital A	Biofilm swab	2012
A39	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A4	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	AQU	Biofilm swab	2012
A40	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A41	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital A	Biofilm swab	2012
A42	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A43	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A44	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital A	Biofilm swab	2012
A45	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A46	6 Dresden	1438	+	Gt16(1)	9,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital A	Biofilm swab	2012
A47	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm	2012
A48	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A49	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A5	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	AQU	Biofilm swab	2012
A50	(2-14)	1438	+	Gt16(1)	9,8,8,2,4,13,2,2,18,3,10,4,8	VACC5	Hospital A	Water sample	2012
A51	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2012
A52	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A53	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2013
A54	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2014
A55	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital A	Biofilm swab	2014
A56	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm	2014

								swab	
A57	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2014
A58	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2014
A59	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2014
A6	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A60	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A61	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A62	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A63	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A64	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A65	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A66	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A67	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital D	Biofilm swab	2012
A68	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital D	Biofilm swab	2012
A69	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital D	Biofilm swab	2012
A7	8	1358	+	Gt11(87)	7,7,10,2,4,17,1,1,18,3,14,5,8		AQU	Biofilm swab	2012
A70	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital D	Biofilm swab	2012
A71	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital C	Biofilm swab	2012
A72	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital C	Biofilm swab	2012
A73	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital C	Biofilm swab	2012
A74	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A75	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A76	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A77	1	1	+	Gt4(16)	7,7,10,2,4,9,4,2,17,2,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A78	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A79	(2-14)	1326	+	Gt13(72)	8,8,11,1,4,16,1,1,3,2,6,4,8	VACC2	Hospital B	Biofilm swab	2012
A8	6	187	+	Gt84(106)	8,8,11,1,0,16,1,1,3,3,6,4,8	VACC2	AQU	Biofilm swab	2012

A80	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A81	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A82	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2013
A83	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A84	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2013
A85	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2013
A86	(2-14)	461	+	Gt9(92)	9,8,11,2,4,13,3,4,17,3,10,4,8	VACC11	Hospital B	Biofilm swab	2012
A87	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A88	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A89	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A9	1 OLDA	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	AQU	Biofilm swab	2012
A90	1	1	+	Gt4(17)	7,7,10,2,4,9,4,2,17,3,14,5,5	VACC1	Hospital B	Biofilm swab	2012
A91	(2-14)	1482	+	Gt8(7)	10,8,8,2,4,13,2,2,18,2,10,4,8	VACC5	Hospital E	Biofilm swab	2014
A92	(2-14)	1482	+	Gt8(7)	10,8,8,2,4,13,2,2,18,2,10,4,8	VACC5	Hospital E	Biofilm swab	2012
A93	(2-14)	93	+	Gt24(68)	8,8,11,2,0,16,1,1,3,0,6,4,8	VACC2	Hospital E	Biofilm	2012
A94	(2-14)	1438	+	Gt16(3)	9,8,8,2,4,13,2,2,18,2,10,5,8	VACC5	Hospital E	Biofilm swab	2012
A95	6 Dresden	1326	+	Gt13(72)	8,8,11,2,4,16,1,1,3,2,6,4,8	VACC2	Hospital E	Biofilm swab	2012
A97	(2-14)	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2012
A98	6 Dresden	1326	+	Gt13(72)	8,8,11,2,4,16,1,1,3,2,6,4,8	VACC2	Hospital F	Biofilm swab	2012
A99	6 Dresden	461	+	Gt10(93)	9,8,6,2,4,13,3,4,10,3,10,4,8	VACC11	Hospital F	Biofilm swab	2013

[1]Serogroup, monoclonal subtype. NA: Not Avaiable. AQU: Al-Quds University

[3] MLVA allele order: *Lpms1, Lpms3. Lpms13, Lpms17, Lpms19, Lpms31, Lpms33, Lpms34, Lpms35, Lpms38, Lpms39, Lpms40, Lpms44*

Curriculum Vitae

Ashraf R. N. Zayed

Date of Birth: May 2nd, 1987

Place of Birth: Nablus, Palestine

Work Experience

Oct 2014-Oct 2017	**Helmholtz Centre for Infection Research**, Braunschweig, Germany PhD Project: *Legionella populations in freshwater systems (DFG)*
Sep2011-Aug 2014	**Al-Quds University**, Abu Dies, Jerusalem, Palestine MSc Project: *Legionella populations in freshwater systems (DFG)*
Sep2011-May2014	**Al-Quds University**, Abu Dies, Jerusalem, Palestine Faculty of Medicine: *Research and Teaching Assistant*

Higher Education

Oct 2009- May 2013	**MSc in Medical Laboratory science/ Microbiology and Immunology** at Al-Quds University, Jerusalem, Palestine
Oct 2005 -May 2009	**BSc in in Medical Laboratory science** at Arab American University, Jenin, Palestine